ERC Synergy Grant 2022
Research proposal [Part B2]

- cPI : Thomas Gregor, PhD (Institut Pasteur, Paris, France)
- PI : Denis Duboule, PhD (Collège de France, Paris, France)
- PI : Gasper Tkacik, PhD (IST Austria, Klosterneuburg, Austria)

Section a. State-of-the-art and objectives

During mammalian embryogenesis, key events involving DNA and regulatory molecules over seconds and nanometers affect, and are affected by, major reorganization of the genetic material in the nucleus over hours and micrometers. How these scales are spanned and integrated into the 4D frame of development remains a major unresolved challenge: regulatory molecules diffuse around and bind to DNA sequences to determine the activity of genes, often in parallel with chromatin rearrangements and the assembly of multi-molecular complexes. However, despite the importance of these processes for our fundamental understanding of how living organisms develop, we currently lack a comprehensive and integrated picture of how the nuclear organization of the DNA and the dynamic interactions between control molecules and chromatin eventually contribute to gene activity and function, and how this can be achieved by bridging such different space and time scales. **Our central goal is to combine nuclear architecture, molecular dynamics and functional activity, in order to provide a dynamic mechanistic picture and a corresponding mathematical model for how various regulatory elements are interacting, how chromatin micro-architectures are formed and maintained as well as to determine the causal relationships between these mechanisms and gene activity along the course of embryonic development.**

The past ten years have seen an explosion of information regarding higher-order loop domains, such as e.g. topologically associating domains (TADs)[1–4], using whole-genome analytical methods such as ChIP-Seq and chromosome conformation approaches. While it is now becoming accepted that TADs reflect widespread units of gene regulation[5], the literature contains conflicting conclusions regarding the causal role of the 3D genome for proper gene regulation[6]. For example, abrogation of particular levels of chromatin organization after removal of some of the factors involved (e.g. the 'architectural' zinc-finger protein CTCF[7]) had a surprisingly low impact on gene expression, at least in the analyzed biological contexts[8–10].

The recent development of sophisticated imaging methods including high-resolution fixed tissue methods and live cell imaging have enabled the first direct visualization of long-range enhancer–promoter interactions[11–14]. These studies uncovered transcriptional bursts, the co-activation of linked genes, and large distances separating enhancers and their target promoters even during transcriptional activation. Such observations challenge our view of transcriptional control and suggest that higher-order associations of regulatory landscapes and associated transcription factors may produce dynamic biomolecular condensates at active genes[15,16]. Yet despite these advances, an **integrative picture of how dynamic changes in chromosome organization control the timing and levels of transcription, and how this timing integrates into a developmental time scale, is still missing**. We postulate that answering this challenge is key to bridging from molecules to tissues and will transform our view of how genes are regulated during mammalian development.

Successful elucidation of the underlying mechanism(s) depends on measuring each critical component of nuclear organization underlying gene activity, including the topological arrangements of cis-regulatory elements (CREs) such as enhancers, promoters and insulators, dynamic long-range chromosomal loops and transcriptional condensates. Hi-C interaction maps generated during the past decade have provided a starting point for such studies[17–19], but they usually fail to reveal the dynamic properties of chromatin organization and gene activity over developmental time, and linking them to transcriptional output in real-time remains challenging. **To achieve this complex task, we must develop analytical paradigms that comply with a number of essential parameters**. The extent and complexity of regulatory landscapes in *Drosophila* is below that observed in vertebrates, where genome duplications allowed for a massive extension of these

landscapes along with a higher level of pleiotropy[20]. Therefore, a developing vertebrate paradigm, preferentially of mammalian origin, should be used, considering the potential impact for the understanding of pathological states triggered by failures in gene regulation, for example following genomic rearrangements[21].

One essential parameter is the choice of model genomic loci to work with. They should be conducive to an extrapolation of potential discoveries genome-wide. Indeed, most experimental approaches require the selection of a few genetic loci, which should ideally represent illustrations of the major regulatory configurations that are found around mammalian developmental genes. **We will therefore concentrate on a few well-defined regulatory configurations**. The first configuration represents the simplest situation, where a unique target gene responds in a given cell type during development to a single strong enhancer sequence located at a distance[22,23]. In the second configuration, several such tissue- or cell-type-specific enhancers, acting on one specific gene, are grouped together to maximize the regulatory input[24]. The third and most challenging regulatory configuration comprises mixed series of tissue-specific enhancers, spanning a large distance and regulating several nearby genes at various times, in various cell-types and with distinct affinities[25].

Another essential parameter is the choice of the model organism, which should comply with unusually demanding criteria. For the sake of comparison, the same organism, at different times and in its distinct tissues, should allow the study of the various regulatory configurations mentioned above, which *de facto* rules out most if not all cellular systems *in vitro*. It should be amenable to direct microscopic visualization, be available in large amounts for molecular studies and be easily accessible for mutations for functional approaches; these criteria disqualify genuine mammalian embryos. To overcome these difficulties, **we propose to use 'gastruloids', which are embryonic stem-cell-derived pseudo-embryos grown in culture**[26,27]. Mutant loci can be easily produced through CRISPR-based approaches and large amounts of biological material are readily available and can be precisely staged along developmental time. When cultivated under appropriate conditions, gastruloids generate several cell- and tissue-types that allow experimental access to the various regulatory configurations mentioned above[28,29]. We thus propose to use these biological objects as test-tubes to study the dynamics of complex mechanisms related to the control of transcription and its relationship with the 3D genome.

The last essential parameter is the choice of language for describing our findings, integrating them, and formulating quantitative predictions. Mathematical modeling of chromosomal configurations from Hi-C snapshots shows that a multitude of configurations are compatible with experimental data in any given biological context[30]. We thus need to think in terms of probability distributions for these configurations[31,32], rather than a single representative one. Similarly, CRE interactions and the resulting bursts of gene activity are stochastic[11], and likely modulated in a probabilistic fashion by local chromosome architecture and associated molecular factors. Therefore, a quantitative understanding of our experiments cannot be formulated as a biological "cartoon" or a simple schematic. It must be embodied in **a mathematical framework that combines the well-understood constraints of polymer physics with the ability to simulate the full dynamics of gene regulatory loci**.

Such a model will provide an essential bridge across scales: spatially, it should predict how mesoscopic chromosome organization affects microscopic events at loci that govern transcription; temporally, it should link continuous series of experimental snapshots of chromosomal organization with measured real-time dynamics of transcribed loci. Deriving such a model is not simply a matter of taking a computational framework that already exists and "fitting" it to data, after the experimental part of the proposal is finished. Instead, formulating and inferring such a model in the complex gastruloid context is a theoretical feat that has not yet been achieved to date. To reach this goal, **we propose an integrated approach between theory, experiments, and modeling**. Initial experiments will be captured by individual models to provide immediately testable predictions that will refine the models. The goal of such iterations is to systematically converge to a multi-scale model that integrates chromosomal architecture with transcriptional activity.

Objectives

Our overarching goal is to integrate three major dynamic viewpoints for a few selected gene loci: the global dynamics of architectural rearrangements in time through contact probability maps, the local dynamics of particular cis-regulatory DNA elements (CREs)[33] involved in shaping locus architecture and gene activity, and the dynamics of the transcriptional output of the locus. The aim of our multidisciplinary approach is to provide a quantitative structure-function relationship that links the spatiotemporal architecture of a mammalian gene locus with its functional output.

The proposed program is articulated around two main phases. **Phase one** starts with a descriptive assessment of the various regulatory mechanisms at work in distinct model loci. We will select these loci based on their potential heuristic value in terms of regulatory strategy and behavior. The experimental data will be analyzed and interpreted through iterative data-driven modeling in each individual task and simultaneously integrated into a multiscale dynamic polymer model, which will inform the **second phase,** where selective perturbations will be introduced into the system.

Phase I objective: To build a quantitative description of chromatin and transcription dynamics

Aim I.1: Hi-C animations, spatial locus architecture, inverse modeling of polymer configurations
Aim I.2: CRE dynamics, gene activity, inverse modeling of effective forces between labelled loci
Aim I.3: Development of a dynamic polymer model of chromatin DNA, inferred from I.1 and I.2

Phase II objective: To challenge quantitative model built in Phase I via functional perturbations

Aim II.1: Genetic perturbation of chromatin landscapes
Aim II.2: Light-induced perturbation of loop formation and stability
Aim II.3: Visualization and perturbation of biomolecular condensates

Section b. Methodology

Phase I

The objective of Phase I is to provide an integrated, quantitative description of chromatin and transcription dynamics. Three aims will reflect this objective: 1) Delivering dynamic gene locus-specific Hi-C maps from gastruloids or parts thereof. 2) Assessing simultaneous *in vivo* CRE interaction dynamics and locus transcription measurements. 3) Providing a general modeling framework that integrates these different data sets. We will define a handful of specific loci in the mammalian genome on which we will perform our experiments and develop both genomic and visualization technologies to obtain unprecedented spatiotemporal resolution. Our modeling framework will eventually incorporate CREs, chromatin interactions between occupied CTCF sites[7], and the dynamics of molecular factors involved in these contacts (e.g. the cohesin subunits and loading factors)[34,35], and also global, less specific interactions occurring at these loci, possibly due to aggregate formation[36–38]. An integration of these aspects will reveal the causal and mechanistic relationships between chromatin structure and transcription, for instance by establishing the importance of precise temporal ordering of molecular events. Also, it will reveal whether all regulatory landscapes operate in the same way, regardless of their intrinsic properties (size, number of enhancers and genes etc.), or whether the regulation of developmental genes relies on fundamentally different types of processes.

Aim I.1: Hi-C animations, spatial locus architecture, inverse modeling of polymer configurations

This aim focuses on the 3D polymer configuration for a selected set of gene loci. A key innovation will be to measure how these configurations change as a function of a developmental time-series and use modeling to infer localized changes on the DNA that drive such rearrangements. We will use several approaches applied to a large series of precisely staged developing gastruloids, to either validate or select the few genomic loci which will form the basis for the proposed research. These approaches will mostly consist of chromosome interaction maps, epigenetic profiling of chromatin marks, DNA binding patterns of particular proteins, and validation of potential enhancer sequences. All datasets will be used to start parallel inference and modeling efforts.

I.1.1 High-resolution 3D profiles of micro-dissected gastruloids and locus selection

Gastruloids will be produced by the aggregation of 300-400 ES cells, followed by a short exposure of 24 hours to the *Wnt* signaling pathway through the GSK-3 inhibitors CHIR99021[27]. Shortly thereafter, cellular movements are detected and an elongation process starts to occur, which resembles embryonic axial elongation. Comparative transcriptomic analyses revealed indeed that the appropriate sets of genes are activated at the expected times and in the appropriate sequence, when compared to normal posterior embryonic axial extension[27,28]. The expected variety of cell types was also scored, indicating that tissue differentiation occurs along with growth and extension. Under our culture conditions, single-cell RNA-seq experiments indicate that gastruloids mostly consist of a pseudo-embryo analogous to the post-occipital part of normal embryos (Fig. 1).

The choice of loci to investigate will be based on several parameters, including strong transcriptional activity during gastruloid development, which should be temporally defined and spatially localized. For example, a key developmental gene that would be robustly transcribed between the third and the seventh experimental timepoints (T3 and T7) in the 'posterior' third of the extending gastruloid would be a candidate. These criteria can be applied to datasets already available in our labs, including single cell RNA-seq and ATAC-seq, allowing us to further map the potential positions and activities of associated enhancer sequences (Fig. 1). Accordingly, we compiled a preliminary, non-exhaustive list that includes *Sox2*, *Pcdh8*, *Cyp26*, *Brachyury,* and *HoxD* loci. In all these cases, preliminary analyses indicate that 1) they are activated at a precise stage of gastruloid development, 2) this activation occurs in a substantial number of cells, 3) the time of activation coincides with specific signal in the ATAC-seq profiles. The *Cyp26a1* locus illustrates our approach and shows an activation in 'posterior-most' cells of the gastruloid at 96h, concomitantly with the appearance of several ATAC-seq peaks (Fig. 1). The restricted expression of this gene (Fig. 1) will allow for targeting optical approaches to the right domain while using neighboring cells as a negative control.

An initial Hi-C experiment will dissect gastruloids at 120h into three compartments, 'anterior', 'middle' and 'posterior' (Fig. 1), such that in the same three Hi-C datasets, all selected loci will be analyzed in both positive and negative transcriptional states. Selected loci will then be kept only if Hi-C maps (ca. 1kb resolution) suggest structural differences upon transcriptional activation. If not, other candidate loci will be examined. In the former case, the potential regulatory function of distinct ATAC-seq peaks will be assessed. When compared with similar H3K27ac profiles, peaks positive for both ATAC and H3K27ac will be analyzed in the transgenic setting, which is already set up in our labs[39],

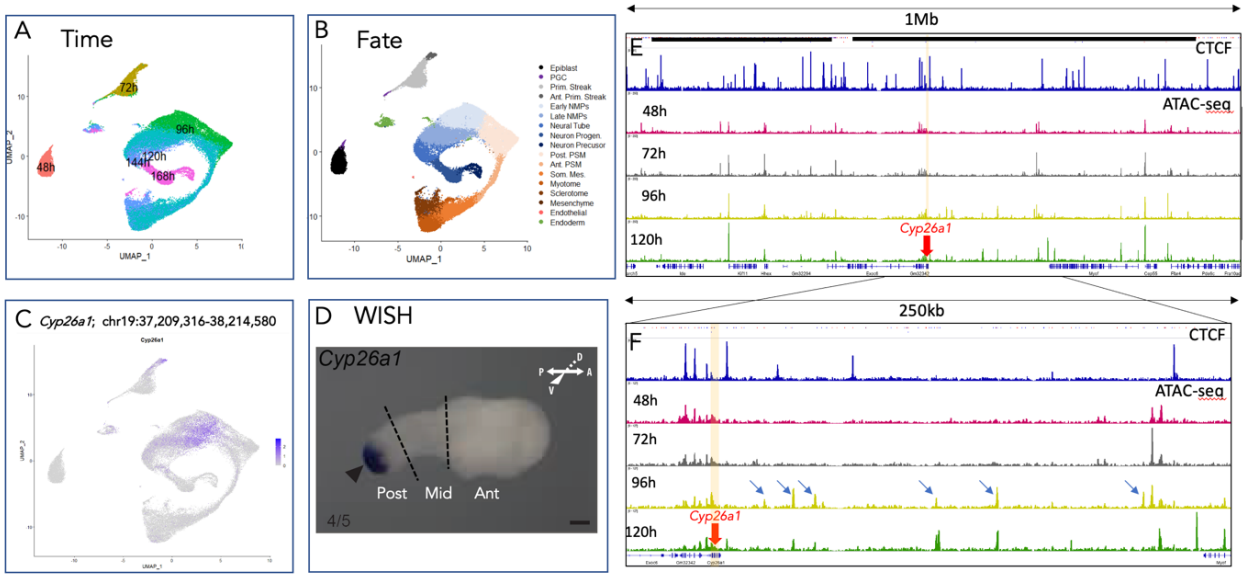

Figure 1. Strategy for the selection of model genomic loci. *Example of a candidate gene locus, preselected before the initial Hi-C profiles on micro-dissected gastruloids. (A) Single cell RNA-seq of gastruloids at four different times. The clusters are shown according to the various timepoints (in colors). (B) The same clusters are marked with cell fates (colors as in the right column). (C) On top of these clusters are shown cells displaying high expression levels of Cyp26a1 RNAs (in blue), indicating their enrichment among NMP cells. (D) WISH shows the expression domain of Cyp26a1 at 120h, corresponding to the clusters shown under (A-C). The dashed lines indicate the dissections used for the initial Hi-C experiment. (E) On top is a CTCF profile and below the ATAC-seq profiles corresponding to the four timepoints displayed in (A). The position of the Cyp26a1 gene is indicated with a red arrow. (D) Enlargement of the profile under (E) showing several ATAC-seq peaks appearing at 96h, i.e., corresponding to the onset of gene activation (arrows).*

and involves a homologous recombination into a mutant *hprt* locus. Recombination frequency will be enhanced after the introduction of a single DNA break through the CRISPR/cas9 system. The selection strategy combined with the high recombination frequency yields a high enough fraction of transgenic cells to carry out this analysis as a medium-throughput, routine approach. Also, recombination at this locus tends to occur in single copy, which prevents artifacts due to multi-copy transgenes and position-effects caused by random integrations in the genome. Only enhancers showing the expected pattern (that of the gene under scrutiny or part thereof) will be further considered for analysis as candidate sequences for DNA tagging and imaging experiments in I.2.

I.1.2 Modeling dynamic high resolution 3D contact maps for selected loci

Once representative loci are selected, their interaction dynamics will be assessed by using a series of precisely staged gastruloids every 8 to 10h, from 48h to 168h (i.e., at 12 to 15 different time points), and reconstructing Hi-C animations to produce dynamic profiles in time. Such a dynamic Hi-C representation has been produced at one locus (Fig. 2) and should be readily feasible for any of our selected loci with dense time sampling and at high resolution. The alternative use of Micro-C[40] (with MNase treatment) may be considered due to the rapid evolution of these technologies[41]. Also, once genetic loci are eventually selected, Capture-Hi-C[42] might be considered, in particular during Phase II where multiple mutants will be produced at the same locus (see below).

These dynamic Hi-C datasets will provide a direct input to maximum-entropy inverse modeling[32,43,44], where Hi-C data are used to infer the underlying interactions between parts of the chromosome, allowing prediction of the distribution of 3D chromosome configurations[45] (Fig. 2). Importantly, when applied to a dense timed sequence of Hi-C snapshots (which has not been done to-date), this methodology will reveal which interactions between chromosomal DNA loci remain constant through time and which interactions must dynamically change to account for the data. Thus, this modeling approach will filter out artefacts due to time- and ensemble-averaging over chromosomal configurations, which preclude direct interpretation of Hi-C contact maps in terms of localized molecular changes on the chromosome.

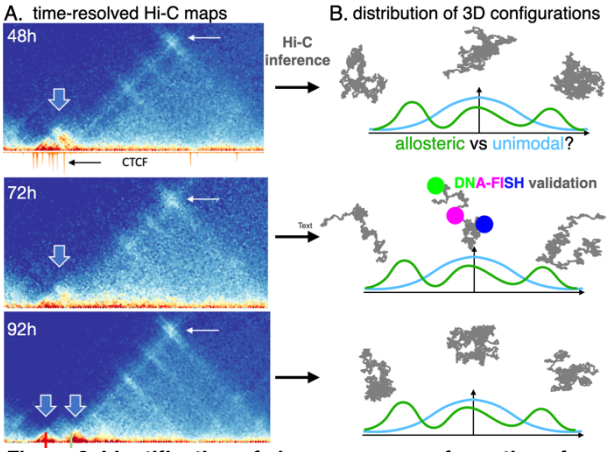

Figure 2. Identification of chromosome conformations from time-resolved contact maps. *(A) Hi-C maps (5kb resolution) as a function of developmental time, with Wnt treatment 48-72h. Green and red bars indicate Hoxd4 and Hoxd9 respectively. Blue arrows indicate the domains. White arrow indicates the anchor point with the first CTCF (black arrow). (B) Concept sketch of how, using Hi-C maximum entropy inference, we will be able to infer the distributions of 3D chromosome configurations consistent with the Hi-C maps as a function of time, validated by independent DNA-FISH experiments.*

I.1.3 Dynamic profiling of molecular players involved in chromatin organization

To generate a comprehensive view of which molecular players drive potential changes in chromatin configurations upon transcriptional activation and termination, we will produce similar time series of datasets for H3K27ac, CTCF, RAD21 (a cohesin subunit) and NIPBL, one of the loading factors of the cohesin complex. Datasets will be handled as described above for the dynamic Hi-C animations, by binning the 'loci' (defined as the TAD containing the transcription unit under scrutiny) to an appropriate size (between 10 to 20bp) and averaging *in silico* all intermediate time points (e.g. see https://s9.gifyu.com/images/54zqw9fw6zrqsdfh0r659.gif). When compared to dynamic Hi-C profiles, these binding profiles assembled over 4 to 5 days of culture will tell us about the relationships between the presence of active chromatin (genes, enhancers), the concomitant recruitment of cohesin (NIPBL) and its accumulation (RAD21) at CTCF sites. This information will indicate where chromatin extrusion[35] occurs and where loops may be formed and stabilized, thus linking the evolution in time of chromatin structures to the behavior of potential drivers of these modifications.

Dynamic profiling will directly connect to the computational output of I.1.2: there, maximum entropy models point to localized temporal changes in the interaction of DNA loci in a way that is completely agnostic to molecular data. Correlations between maximum-entropy-identified changes and dynamic profiling from I.1.3 will generate causal hypotheses about how molecular marks affect structural changes. In return, refined maximum entropy models can be inferred and tested by assuming that

interactions driving 3D chromosome configuration changes can only occur where significant changes in dynamic profiling data have been detected. This defines a theory-experiment feedback loop that identifies the minimal set of necessary molecular marks required to explain chromosome configuration changes, consequently informing on the nature of perturbations to be induced during Phase II.

Aim I.2: CRE dynamics, gene activity, inverse modeling of effective inter-locus forces

The genome topology at most gene loci is too complex for a purely static approach to reveal structural and functional control of gene activity. Therefore, we will complement I.1 with a dynamic approach based on live imaging of gene transcription, to dissect their bursting kinetics and their spatiotemporal regulation at the cellular and the single-locus scales. We will establish the well-described MS2 and PP7 stem-loop arrays[46–49] in gastruloids to label nascent mRNA of multiple genes at the selected loci, and will implement the orthogonal Anchor1 and Anchor3 systems for DNA tagging *in vivo*[50,51]. Multiple DNA elements will thus be followed in real time to provide insights into chromatin dynamics and the interactions between enhancers and promoters, as well as into their structure-function relationship. For control experiments, we will also establish imaging based on fixed gastruloids (single-molecule RNA FISH[52,53] and an adapted multicolor oligopaint approach[54–56]). This will help with screening genes rapidly and serve as quantitative calibration. DNA-FISH-based intra-locus CRE-distance measurements will also bridge between our imaging experiments and the dynamic locus rearrangements observed and modeled in the Hi-C profile series from I.1.

I.2.1 Establish gastruloids as a quantitative model for *in vivo* fluorescence imaging

An important preliminary step will be to establish gastruloids as a quantitative model for *in vivo* fluorescence imaging. The Gregor laboratory has successfully done this with the early fly embryo, but a number of technical challenges need to be resolved to turn the larger, less stable and slightly less reproducible gastruloid structures into a similarly productive quantitative assay.

Construction of ES cell lines with transcription readouts: A series of ES cell lines will be produced using CRISPR-Cas9-based homologous recombination to target our selected loci. Destabilized fluorescent marker sequences will be introduced to visualize live the expression of the gene product and to sort out positive cells for further analyses, an approach already established in the Duboule laboratory. To monitor transcription dynamics in single cells, we will implement the stem-loop-based MS2 and PP7 fluorescence reporter systems[46–49], which allow for the tracking of newly synthesized RNA by tagging endogenous gene introns. This will allow us to precisely monitor transcriptional dynamics with single molecule resolution and to identify the activity status of particular cells with respect to the ongoing transcription of the selected target gene. Used in conjunction with spectrally separable fluorescent tags it will also provide the necessary insights for colocalization and coactivation of multiple genes located either in *trans* or in *cis* configuration at the same locus.

While MS2/PP7 labeling systems can be used to simultaneously track the location of active genes in the nucleus, it is often just as important to track inactive genes and hence transcription-independent labeling systems are needed. We will use the recently developed parABS DNA labeling system, which employs Burkholderia parS DNA sequences that nucleate the binding of a ParB-GFP fusion proteins[57]. This approach is thought to be less disruptive to the local chromatin structure than more traditional DNA labeling systems such as TetO/TetR or CuO/CymR[58]. We propose to use the ANCHOR1 and ANCHOR3 implementations of this approach that have been successfully implemented in mESCs[51]. However, while successfully implemented in *Drosophila* embryos carrying multiple transgenes, these imaging approaches have yet to be reported for other biological objects. We thus anticipate that some R&D will be necessary (already started in the TG and DD laboratories).

Develop rapid and quantitative long-term gastruloid imaging modes: Several imaging modalities will be used to perform the imaging assays proposed in this grant, such as highly inclined and laminated optical sheet (HILO) microscopy[59], Bessel beam selective plane illumination using two-photon excitation[60,61] and oblique plane microscopy[62] (OPM). The break-through advantage of OPM is that when implemented with remote focusing[63], a single high numerical aperture objective is used for both selective plane illumination and emitted fluorescence detection, overcoming the typical limitations of standard selective plane illumination, thereby achieving subcellular resolution with

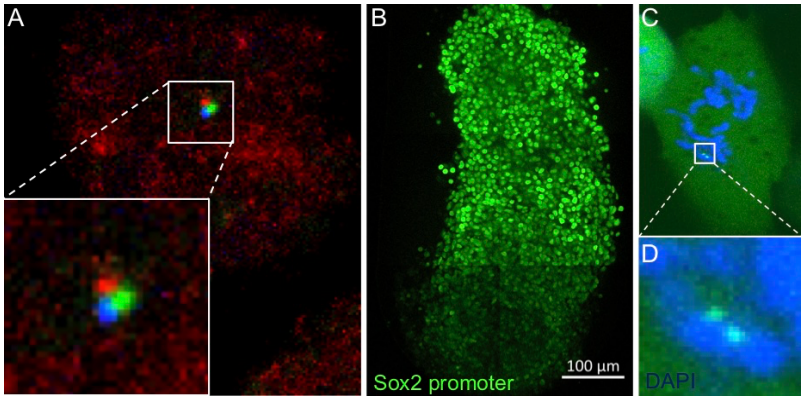

Figure 3: Imaging enhancer–promoter interactions in live mESCs and developing gastruloids. ***(A)*** *An mESC expressing Sox2-PP7 bound by PCP-BFP (blue), with parS sites at the endogenous Sox2 promoter (bound by ParB-eGFP, green), and a tetO cassette at the Sox2-SCR (enhancer), bound by TetR-mKate (red).* ***(B)*** *2-photon imaging of 120h old gastruloid with labeled Sox2 promoter.* ***(C)*** *Close-up of cell from B with chromosomes (blue).* ***(D)*** *Two paralog Sox2 promoters on sister chromatids.*

single-molecule sensitivity. Recently, this method was further developed to achieve volumetric imaging of living samples at high speed up to 300 volumes per second[64,65] and to expand the usable effective numerical aperture of the objective[66,67]. In this way it is possible to visualize and quantify molecular dynamics and interactions in both living cells and gastruloids. By employing a wide array of different imaging modalities, we will tailor the custom-built instruments (available in TG laboratory) to our specific applications, thereby minimizing risk and managing fallback solutions.

The development of quantitative imaging of transcription events and chromatin dynamics in the fly embryo was successful largely due to their intrinsic behaviors under the microscope[68]. They are virtually immobile and resistant to environmental fluctuations or flow of medium. In contrast, gastruloids are exposed to convective flow and are thus poorly amenable for high-quality imaging experiments. We will need to develop several technologies to generate reproducible imaging conditions, both in custom-designed microfluidic environments that can house developing gastruloids for long-term culture and simultaneous imaging, as well as in gel-like embedding media of various densities that allow for continuous flow of medium and natural gastruloid elongation, while further stabilizing the specimens inside the growth chambers.

I.2.2 Multi-color imaging of CREs and transcription in cell culture and gastruloids

As a first step in applying our imaging tools to gastruloids, we analyzed the *Sox2* locus in cultured mESCs. We selected this locus since its expression is necessary to maintain pluripotency in mESCs and it can be clearly detected in gastruloids throughout their development. The *Sox2* RNAs also mark differentiating neuronal precursor cells, which develop after 120h in our gastruloids. As a pilot experiment, we can monitor *Sox2* gene activity in both systems and determine the relationship between the *Sox2* gene body and its essential enhancer, the *Sox2* Control Region (SCR) located 100kb far from the gene itself[14,69]. We are implementing our imaging assays at the *Sox2* locus (Fig. 3) in combination with the ANCHOR1/3 localization tags at the distal enhancer and promoter, along with the MS2/MCP system to visualize transcription (e.g., see system implemented in fly embryo: https://s10.gifyu.com/images/3M8QT8UVRZRYF44TE4G.gif). Preliminary experiments using two-photon microscopy show promising results, with the *Sox2* promoter identifiable in individual living cells in developing gastruloids (Fig. 3). As part of this approach, we will engineer tagged SCRs at different distances along the chromosome. This will enable us to systematically study the effects of enhancer-promoter distance on gene activity[70]; it will provide an important test of the polymer model proposed in I.3.3. Once established, this approach will be applied to the other selected loci.

As an initial paradigm to study genes located in *cis* in the same TAD and activated following a time sequence, we will investigate the transcription dynamics for pairs of *Hox* genes separated (or not) by an occupied CTCF site[25]. This will allow us to assess whether loop domains coordinate transcriptional bursting in time for genes linked in *cis*. We will focus on the impact of defined enhancer elements that control *Hox* gene expression during their early phase of activation, which can be precisely observed in staged gastruloids. To concomitantly measure gene activity for two *Hox* genes linked in *cis*, we will use CRISPR/Cas9 to insert MS2 and PP7 stem loops in these genes (e.g., the *Hoxd4* intron for MS2, and the intron of either *Hoxd1* or *Hoxd8* for PP7), thus allowing us to verify that transcription, rather than mRNA accumulation, indeed occurs in a time sequence within each cell, something that remains to be documented. To produce these cell lines, as well as to prepare ES cells for experiments in Phase II, we will start by generating mESCs lacking one full copy

of the corresponding region, including the two TADs as well as the entire *HoxD* cluster (to produce a haploid locus). Thus, all genetic manipulations will be performed on the same chromosome, thereby preventing the production of homozygous configurations. This will make future genetic manipulations more convenient, both for CRISPR/Cas9 modifications as well as for screening colonies, as only one allele will be targeted. Such haploid configurations will be produced for other loci selected for genetic perturbations in Phase II.

I.2.3 Force-inference modeling of physical forces acting on CREs

Real-time tracking experiments will rapidly provide data for "force-inference" inverse modeling, where the trajectories of thousands of tagged loci in individual nuclei are used to reconstruct effective physical forces acting on these loci due to the local chromatin context. Several methods for inferring forces from stochastic trajectories have been proposed in recent years[71–73], but they have not been applied in the context of chromosome dynamics. This modeling will enable us to infer whether effective forces change prior to, or concomitantly with, the change in transcriptional status of the locus, thereby revealing physical mechanisms that bring promoters and enhancers into productive contacts or sufficient proximity to drive transcriptional activation. Importantly, while the distance between the loci undergoes large random fluctuations from moment to moment due to molecular and thermal noise, the inferred force will filter out this stochasticity and will only change when the underlying biological mechanisms change. Simultaneously, the inferred force will be compared with the transcriptional bursting activity of a given locus. A key question here is whether or not a separation of timescales exists. On the one hand, changes in the force could be slow compared to bursting kinetics, suggesting that bursting is more (less) likely, or may happen with a higher (lower) frequency depending on the force that acts between the promoter or enhancer[74,75]. On the other hand, the force could fluctuate on a timescale that is comparable to expression kinetics, suggesting that each burst is directly enslaved to a fluctuation in the force. Aim I.2 would thus help resolve this essential question and inform the integrated model of I.3.

Force-inference modeling also has major implications for our understanding of DNA polymer dynamics, a major component of the integrated model of I.3. Current theory can explain the dynamics of single chromosomal loci in terms of sub-diffusive motion in the viscoelastic nucleoplasm. Theoretical predictions have been worked out for the joint motion of two loci based on simple polymer models[76,77], but have not been tested experimentally. To test these predictions, we will extract dynamic correlations (e.g. two-point correlation functions or velocity cross-correlations), giving access to the dynamics of stress propagation (shown in Fig. 4 as a proof-of-concept for *Drosophila* data). Importantly, we expect the theoretical predictions to apply to joint statistics of arbitrary sites; sites with specific interactions such as enhancers and promoters may require development of new theory. We will thus initially test these predictions on experiments with arbitrary tagged DNA loci as a stepping-stone towards studying genuine enhancer-promoter interactions. We will then connect the observed correlations to the observed chromosome configurations based on Hi-C (Aim I.1), by extracting scaling exponents related to the spatial compaction of the chromosome, building on recently developed theory[76]. Using state-of-the-art inference methods[78], we will then perform force inference (see Fig. 4 for a proof-of-

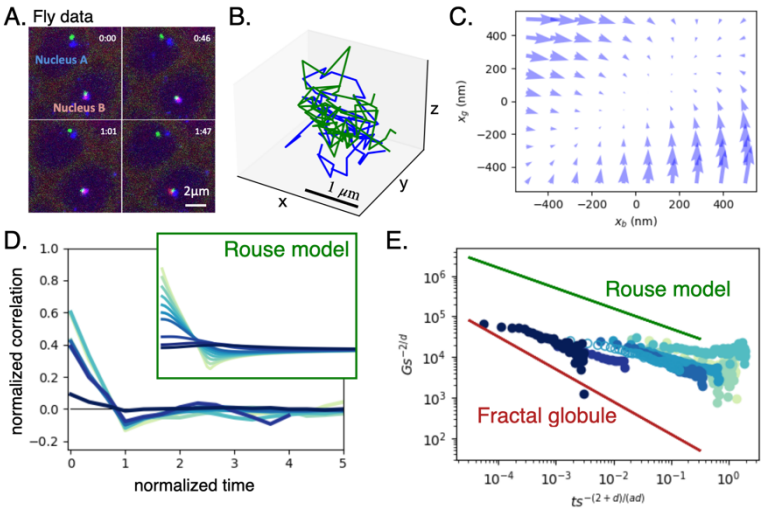

Figure 4. Force inference and model testing with Drosophila data. *(A) Time series snapshots of two nuclei with labelled eve promoter (green), enhancer (blue) and transcriptional activity (red). (B) 4D-trajectories of the eve-enhancer and ectopic promoter inserted at a genomic distance s=142kb. (C) Proof-of-principle of an inferred force field obtained by applying Stochastic Force Inference to the trajectories in B. (D) Velocity cross-correlation of the two loci for varying genomic distance. Inset: Prediction for the velocity cross-correlation from a basic polymer model (Rouse). (E) Two-point correlation function of the distance between the two loci, providing an estimate of stress propagation. Predicted power-law exponents for the fractal globule and the Rouse model based on scaling theory are indicated.*

principle) to determine the effective interactions between chromosomal loci mediated by the chromatin and other factors of the nucleoplasm.

Aim I.3: Development of a dynamic polymer model of chromatin DNA, inferred from I.1 & I.2

In Aims I.1 and I.2., we propose to develop two different combined experimental and modeling approaches. In I.1, we extract structural snapshots from large-scale Hi-C experiments and develop inverse models based on maximum entropy. In I.2, we use microscopic real-time imaging experiments to extract the polymer dynamics of DNA monomers on short timescales and develop inverse models based on force inference. The overarching goal for Phase I is to integrate these two very different experimental and modeling approaches into a dynamical polymer model that simultaneously bridges the spatial and temporal scales addressed by these diverse experiments. Aim I.3 represents an entirely novel theoretical challenge, with potential impact beyond the scope of this grant. This is because previous work focused either on reproducing polymer configurations from Hi-C data while neglecting the dynamics, or on establishing fully dynamical models using molecular dynamics (MD) simulations whose effective parameters are matched to sequencing data taken at a single time point. The novelty here is the inclusion of real-time tracking of individual DNA monomers whose statistics the model needs to reproduce, as well as the temporally evolving structure of chromosomal configurations of I.1.2 driven by molecular changes probed by sequencing in I.1.3. The underlying assumption of our model is that these changes evolve on a much slower timescale than the timescale on which the polymer samples its feasible configurations.

I.3.1 Simulations of DNA polymer dynamics with time-modulated molecular markers

We will start with a mechanistic polymer simulation using MD that can be executed in a standard high-performance computing environment using open-source libraries such as LAMMPS or HOOMD[79,80]. An appropriate level of coarse graining represents the DNA as a chain of “beads”, each representing typically ~1kb piece of DNA, whose positions evolve according to Newtonian laws of motion under the influence of backbone constraints, thermal random forces (Langevin dynamics) and specific interactions between the beads due to active molecular mechanisms. As reported previously[81], these mechanisms may include chromatin-protein-chromatin (CPC) interactions and loop extrusion (LE), implemented by polymer sites that preferentially interact, identified using sequencing data (ChiP-seq and ATAC-seq). First, we will extend this approach by allowing CPC and LE interactions to change with developmental time. Second, we will include the ability to explicitly track marked loci, in particular the CREs (promoters and enhancers), and compute joint statistics of their motion (e.g., distance distribution, velocity correlation functions)[76,77]. Third, we will extend the framework to simulate postulated parametrized rules by which these CREs, based on their proximity, lead to transcriptional activity, for instance, by modulating the frequency of transcriptional bursting[75]. Taken together, this will result in a flexible mechanistic model and a technical high-performance computing platform for forward simulation of DNA loci and transcriptional activity, based on a set of mechanistic parameters yet to be identified.

I.3.2 Integration of genomic and imaging data to constrain model parameters

The parameters governing MD simulations proposed under I.3.1 need to be constrained by data. Models built under I.1 and I.2 provide strong constraints on those parameters, while also serving as internal consistency checks between different datasets and models. First, inverse modeling of chromosomal configurations in I.1.3 *predicts* the distance distributions between any pair of chromosomal DNA loci at each developmental timepoint, while this distribution can be directly *measured* by real-time tracking for specific tagged CREs or by *in situ* staining of DNA loci of interest (foci). For each pair of foci, their distance distribution *inferred* by maximum entropy modeling *constrains* the parameters of the forward integrative mechanistic model of I.3.1. In other words, the parameters can be fit to match the inferred constraints.

Second, the inverse modeling in I.1.2 *infers* changes in interaction energies for pairs of DNA loci which should *predict* temporal changes in chromatin mark profiles measured in I.1.3. Together, these results *constrain* at what time points the interaction parameters of the forward integrative mechanistic model of I.3.1 need to change to capture the observed changes in the chromosome structure. Finally, inverse modeling in I.2.4 *infers* the force between foci at different time points. In the forward mechanistic model of I.3.1 this force can be *predicted* given parameters and temporal changes of

the binding factor profiles. Previous theoretical work suggests the statistics of motion of individual foci on the polymer, as well as the statistics of correlated motion between pairs of foci, as diagnostic about the nature of the polymer. We will use the real time measurements of both arbitrary tagged foci (as proposed in I.2.3) and sites with potential specific interactions (CREs) to *constrain* the parameters of our polymer model.

Taken together, we hereby integrate the information extracted from inverse modeling performed directly on experimental data using methods that make minimal assumptions, with the parameters of a dynamical, mechanistic polymer forward model that need to be fitted. To fit these parameters, we will use standard stochastic large scale optimization techniques[82,83]. This approach to inference suggests a productive theory-experiment feedback loop. Given candidate parameters inferred from the data and using postulated mechanisms (e.g., CPC, LE, and various extensions such as condensates explored in Phase II), our model can predict distance distributions and motion statistics for any pair of loci on the DNA; consequently, it can identify those pairs of locations on the DNA whose distance or motion would change most in response to activation or inactivation of specific mechanisms. This is a direct model prediction for which DNA loci should be targeted in the experiments for tracking or staining to most effectively test postulated mechanisms, and to iteratively constrain the model parameters.

I.3.3 Identify structural and dynamical motifs controlling transcription

Having developed and constrained our mechanistic dynamic polymer model in I.3.1 and I.3.2, we will use this model to identify how biophysical chromosome organization mechanisms control transcriptional dynamics. Specifically, we will position tagged CRE loci (enhancers and the promoter) onto the simulated polymer and identify the typical 3D chromosome structures and the typical dynamics displayed by these CREs during various steps in transcriptional activation. Comparing the conformation ensemble and CRE trajectories when transcription is active *versus* when it is not will help us identify recurring structural and dynamical “motifs” that trigger, are necessary for, or correlate with, transcriptional activity. Using the simulation model, we will then investigate which underlying chromosome organization mechanisms (e.g., CPC, LE, or further mechanisms suggested by experiments in Phase II) give rise to these motifs. Simulated perturbation of these mechanisms can lead to a disruption in the occurrence of the identified motifs, and thus to a predicted change in transcriptional activity if the motifs and mechanisms are causal.

For example, an important determinant in any polymer-based model of interactions between any two DNA loci, in particular for CREs, is the genomic distance separating the two DNA loci. Our model will predict how changing this distance will affect the structural and dynamical motifs and, consequently, transcription. As suggested in I.2.2, these predictions can be challenged in experiments where the SCR is inserted at different locations along the chromosome, providing a strong test of the model. In a similar spirit, simulations will provide predictions for perturbation experiments of various molecular organization mechanisms, which we will explore in Phase II.

In addition to general organizational mechanisms that act on the chromatin overall and thus influence the distance and motion of arbitrary pairs of tracked foci on the DNA[84], there could also be additional specific interactions between the CREs, such as promoter and enhancer(s). To identify putative specific interactions, our model will provide a quantitative null expectation for the promoter-enhancer motion, which can be compared to experimental data. If there is no comparison mismatch, then enhancer-promoter contacts can be fully explained by general chromosomal organization mechanisms. If there are mismatches particular to CREs, they must arise due to *specific* promoter-enhancer interactions that cannot be explained by mesoscopic polymer organization alone. Depending on the nature of the mismatches, we will, in Phase II, postulate and test effective models of how promoter and enhancer might mechanically interact beyond the null expectation.

Phase II

Once a basic dynamic model will have been established during Phase I for the few selected gene loci, it will be experimentally challenged during the second phase of the grant. We will test the functional consequences of genome modifications (such as CRE deletion/disruption) on transcription dynamics to probe causal relationships between genome organization and gene activity. To this end,

our dynamic polymer model will predict which of these functional components are most essential for specific structural features and transcription[85]. Permanent global perturbations will be achieved through the generation of modified ESC lines, where small and precise sequence-level modifications of regulatory elements will be carried out on a haploid configuration (see above). In addition, we will develop assays for acute localized perturbations by establishing optogenetic control through light-stimulation in gastruloids. We will disrupt defined long-range chromosomal contacts using an engineered, photo-cleavable form of cohesin and by stimulated depletion of CTCF. In parallel we will try to interfere with the presence of biomolecular condensates at defined genetic loci. The resulting transcriptional responses of these manipulations will be visualized in gastruloids using the high-resolution imaging methods described above, and analyzed on a whole-genome scale using various HI-C assays as in Phase I. Here, the predictions of our dynamic polymer model will guide the specific measurements required to capture the expected changes in chromosome structure and dynamics. By combining perturbations and imaging tools, we aim to translate the typically static genome maps into dynamic patterns of gene activity in living mammalian cells and tissues.

Aim II.1: Genetic perturbation of chromatin landscapes

To further determine the functions of the CRE elements identified in Phase I, we will use the CRISPR/Cas9 toolbox to induce series of genetic modifications such as deletions, additions, inversions or modifications in relative distances between various CREs. All such modifications will be motivated by the datasets obtained in Phase I. Our model will then predict the importance of such modifications, by simulating how transcriptional activity is influenced by the distance between elements, the presence or absence of CTCF binding sites or tethering elements. All such predictions can be readily assessed by using the prepared haploid cell lines described above.

II.2.1 CRISPR/Cas9 modifications of CREs in selected loci

By taking the *Cyp26a1* locus described in Fig. 1 as a strategic example, the questions addressed under this aim will be how many putative regulatory sequences are necessary to properly activate the target genes at the right time in all posterior cells. Also, are the intervening distances (over the 250kb landscape) important, for instance to provide the space required for the formation of a regulatory structure (condensate?). This will be assessed by deleting the landscape and replacing it by various synthetic pieces of DNA through homologous recombination, containing any distribution of enhancer sequences organized over a short DNA segment. While such experiments are starting to be carried out in various contexts, our gastruloid setup and the work carried out under Phase I will allow for an unprecedented degree of developmental timing and spatial resolution (nascent transcription, tagged DNA sequences), for a gene activated in a restricted cell population with an adjacent cell population for negative control or located nearby.

Analogous regulatory surgery will be carried out on the few other selected genetic loci, depending on the outcome of Phase I. In the case of the *HoxD* gene cluster, the insertion of the MS2 and PP7 reporter systems into two genes located in *cis* should allow us to identify the mechanisms that control the time progression in the activation of this series of genes[86], by inducing various perturbations that, according to the model, may affect the transcriptional timing. For instance, by monitoring transcriptional bursts of multiple genes simultaneously, we should be able to assess whether promoter competition alters bursting frequency. The bursting dynamic observed after the selective removal or addition of neighboring CTCF sites[25], or the inversion of their polarity[87], should inform on the relationships between the precise loop topology and transcriptional control. Variations in the number and topological organization of enhancers in relation with the looping pattern will also be considered, depending on results obtained in Phase I.

Aim II.2: Light-induced perturbation of loop formation and stability

While II.1 is concerned with subtle modifications in the topologies of regulatory landscapes, II.2 takes the opposite angle whereby localized and controlled perturbations will be induced in the elements that organize these regulatory landscapes in 4D. Specifically, we will develop optogenetic tools to acutely disrupt two components of the core of current models of transcriptional regulation, yet in a situation where a direct control can be analyzed in the same sample. The first approach will involve the targeted destabilization of the cohesin complex, while the second approach is concerned with the expression of a light-controllable CTCF protein that can be rapidly depleted from the nucleus.

II.2.1 Development of a photocleavable cohesin complex

A prevailing model for long range gene regulation posits that the chromatin fiber is organized into loops that might bring enhancers and promoters into close proximity, sometimes skipping over genes in the process. This model requires specific binding of factors to insulator/tethering sites to induce the formation of site-specific chromatin loops, followed by the formation of stable mechanical linkages that maintain this 3D genome organization over the required time[88,89]. The cohesin complex, which can encircle multiple DNA strands, is thought to be an essential factor for extruding and maintaining stable DNA loops for 3D genome organization and transcriptional regulation[90]. We thus propose an optogenetic strategy to locally disrupt cohesin linkages in interphase cells while leaving mitotic progression unaffected. We will apply these tools at our selected loci in mESCs and gastruloids to dissect genome organization using the experimental tools developed under Phase I.

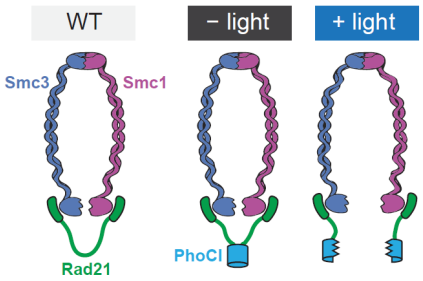

Figure 5: Engineering a light-controllable cohesin complex. *Cohesin is a ring-shaped complex containing two Smc subunits and a kleisin protein (e.g. RAD21) that closes the ring. Separase cleavage of RAD21 normally opens cohesin during mitosis, and prior work with an engineered TEV-cleavable RAD21 demonstrates that inducible cleavage is sufficient to disrupt cohesin's transcriptional functions in interphase cells as well as its roles in mitosis. Here, we will engineer light-regulated kleisin closure. Local illumination will then be used to disrupt the cohesin complex in in a spatio-temporally controlled manner.*

Our approach builds on prior work where a TEV cleavage site was introduced into the RAD21 subunit[91], a necessary component of the cohesin ring complex, fully rescuing the loss of endogenous RAD21. Upon TEV addition, it generates a complete loss-of-function phenotype including an overall mitotic arrest that is expected upon loss of cohesin activity. We propose to replace the slow, genetic induction of TEV expression with a fast optogenetic protein cleavage (Fig. 5). To this end we will introduce a photocleavable fluorescent protein (PhoCl) into RAD21[92] or, alternatively, combine the existing TEV-cleavable RAD21 with optimized variants of a recently-developed light-controlled TEV protease[93]. The use of light in these two approaches should provide direct visualization of cells with or without the modified component, thus readily assessing the effect by employing our reporter systems from Phase I.

This methodology will require some troubleshooting and the possibility exists of leaky activity in the dark or incomplete activity upon illumination. It is therefore crucial to consider available alternative strategies such as the insertion of multiple PhoCl domains into a single RAD21 subunit to increase cleavage efficiency or replacing PhoCl with reversible light-dissociable domains[94]. Perturbation of the *Sox2*-SCR interaction will be an efficient test for this optogenetic technique. Indeed, proximity between the SCR enhancer and *Sox2* depends on CTCF/cohesin-mediated chromatin loops that are cell-type-specific[69]. We will use the triple-reporter system (Fig. 3) as a parental cell line, and will introduce a transgene expressing our light-cleavable cohesin under the control of a strong and ubiquitous promoter. The endogenous *Rad21* gene will be inactivated with a single CRISPR RNA guide such that surviving growing cells should be susceptible to light-dependent depletion of cohesin rings and hence abrogate *Sox2* expression. Once successfully tested, these imaging and perturbation approaches in mESCs will be transferred to gastruloids.

Applying light-cleavable cohesin to pseudo-embryos: By using staged gastruloids, we expect the development of a light-cleavable cohesin ring to help address key questions related to the dynamics of transcriptional control. Light stimulation enables an instantaneous, potent effect by protein photo-conversion, rather than heat shock induction of TEV protease, which only occurs gradually over time. Therefore, we will be in a position to precisely assess how the loss of cohesin-dependent loop formation alters chromosome organization right before, during, or after loop formation, as judged by our time-series datasets obtained under Phase I. In this scheme, entire gastruloids (in case of global interaction readouts such as Hi-C) or individual gastruloid cellular domains (in case of optical readouts) will be exposed to light stimulation at the requested time, depending on the loci under scrutiny. In the former case, Hi-C will be used to obtain a global view of the effect on chromosome architecture of perturbing cohesin function during a precise time period.

In the latter case, by measuring promoter-to-enhancer distances and transcriptional output using the localization-microscopy approaches described under Phase I, we will dissect transcription dynamics either by preventing stable loop formation or by blocking enhancer-promoter localization and transcription. Should cohesin be required for maintenance of enhancer–promoter contacts, its disruption should drive acute loss of gene expression in the corresponding regions and increase enhancer–promoter distance. Conversely, we may find that the liquid-phase condensates driving active transcription also provide an effective surface tension, maintaining a fixed distance between the enhancer–promoter pair and sustaining transcription, even after cohesin linkages are disrupted.

II.2.2 Development of photo-depletable CTCF for rapid conditional inactivation

CTCF is one of the major regulators of chromatin architecture[7]. Along with the cohesin complex, CTCF triggers the formation and stabilization of chromatin loops[35], thereby influencing gene transcription. CTCF is thought to play a major role both as an insulator element, i.e., a DNA segment that prevents the interaction between enhancers and promoters and thus has a negative effect upon transcription, and as an anchoring factor necessary to bring an enhancer close to its target promoter, thus promoting gene expression. The function of CTCF in these contexts has been studied either by deleting binding sites[25,95,96], or by depleting CTCF from cells or embryos[8,10], with different results regarding its global effect on the general chromosome architecture or its localized effect in particular gene loci. We would like to evaluate the function of CTCF by using the analytical tools developed in Phase I, in a system that permits a precise ablation of CTCF function at the various time points of importance for transcription at our selected gene loci, for example just before or after a particular chromatin loop has been established.

In this aim, we will engineer cell lines expressing a light-controllable CTCF protein that can be rapidly depleted from the nucleus. We will fuse CTCF to LEXY, an optogenetic tool based on the AsLOV2 protein domain whose nuclear export can be rapidly and reversibly triggered by blue light[97–99]. Blue light illumination uncages a buried nuclear export sequence (NES) in LEXY's C-terminal Jα helix; in the dark, NES activity is lost and CTCF's nuclear localization signal (NLS) returns the fusion protein into the nucleus. LEXY-based translocation typically produces a 5-fold change in nuclear protein concentration[97,99]. A custom-built 2-photon microscope in Gregor's laboratory has been designed to solve an important problem in developmental optogenetics[100], i.e., how to reconcile the abundance of GFP-channel reporters with precise control over LOV-based optogenetic tools[101] (e.g., LEXY and iLID). It appears that 970 nm imaging enables excitation of, e.g., GFP and mCherry without undesired optogenetic stimulation and can be combined with a digital micromirror device to perform high resolution imaging and precise optogenetic stimulation.

Aim II.3: Visualization and perturbation of biomolecular condensates

Advances in our ability to assess the higher-order arrangement of chromatin in the eukaryotic nucleus have added yet another dimension to our understanding of gene control. This understanding calls for a transition from a classical sequence-based linear view to a gene locus-based spatial view where the three-dimensional arrangement of chromatin produces micro-environments around which the entire locus might be organized[102–104]. These micro-environments can host various factors of the transcriptional machinery[16,105] and have been postulated to serve as hubs for transcriptional biomolecular condensates where enhancers and promoters interact potentially over large (hundreds of nm) physical distances[14,16,106,107]. These distances can potentially explain observations of long-range interactions between enhancers and promoters that do not perfectly colocalize, implying an absence of direct molecular contacts. This view can be reconciled with experimental work in the mouse where the deletion of several enhancer sequences in the same regulatory landscape did not produce a significant effect until a sufficient number of sequences were removed[e.g.,36]. Several candidate loci mentioned under Phase I display an apparent regulatory organization suggestive of such a process (e.g., the *Cyp26a1* locus, Fig. 1).

II.3.1 Site-specific characterization & disassembly of endogenous transcription condensates

We propose to precisely characterize the spatial configuration of such micro-environments, as produced by the selected loci from Phase I, and to design and test an approach to locally inhibit their formation and test the consequences on the transcriptional output. Spatial configurations will be assessed by high-resolution oligo-paint measurements using localization-based microscopy[54,56,107],

where individual elements of the locus will be tagged simultaneously with several multi-colored 1kb fluorescent DNA probes. Disassembly will be attempted by a direct test of the condensate hypothesis, through optical interference with endogenous formation of transcriptional condensates. Currently, no tools are available to either disassemble or interrupt the formation of such endogenous transcriptional condensates. The approach we propose to develop relies on an optogenetic tool to locally block the nucleation or growth of protein condensates at a given genomic locus and at a selected time. It is based on locally concentrating a fusion protein made of an intrinsically disordered protein sequence and the protein purification tag maltose binding protein[108] (MBP). MBP has unique and potent anti-phase-separation properties and when it is fused to the intrinsically-disordered protein region (IDR) from Laf-1 (the so-called RGG domain), the fusion protein blocks RGG liquid-liquid phase separation both *in vitro* and in eukaryotic cells[109]. We propose to extend MBP-IDR fusion proteins to arrest nucleation and/or growth of endogenous transcriptional complexes in mammalian cell lines where engineered protein variants can be rapidly expressed and tested.

We will try to locally deliver MBP to those selected genomic loci where condensates could be observed, while monitoring their formation and the cell's transcriptional response. The rationale is that a high local density of MBP will decrease the local propensity of mediator and transcription factor condensation in the vicinity of a desired genomic locus. To accomplish local MBP targeting, we will use the CasDrop system to recruit 24 IDR-MBP-ferritin subunits per copy of dCas9 to a specific genomic locus of interest[110]. Our long-term goal is to transfer the MBP-CasDrop system into gastruloids and monitor changes in enhancer-promoter distances using the various imaging approaches described above. This study should provide a test as to whether transcriptional condensate formation is directly related to locus compaction or expansion, and whether these are essential for communication across the recently observed ~200 nm gap separating enhancers and promoters[14,106,107]. As an experimental readout, we will again take advantage of our dynamic polymer model. Measurement of locus dynamics and transcriptional activity may not suffice to make a direct causal link to long-range interactions through condensates. However, our mechanistic model might enable us to disentangle the basic polymer interactions from the additional, specific interactions mediated through the condensate, and predict the consequences for transcriptional activity.

Overall feasibility and risk assessment

We hereby propose a high-risk research program with work items of different expected feasibility. Most experiments proposed under **Phase I** should be feasible within the planned timescale given the expertise of the participating groups, including the initial Hi-C experiments on micro-dissected gastruloids, and the precise temporal analyses of chromosome conformations and epigenetic landscapes. Likewise, the implementation of the oligopaint and DNA tagging technologies with optical measurement of distances and gene activity are unlikely to present major hurdles. However, the implementation of MS2 and PP7 systems to detect nascent transcripts in a mammalian context, *a fortiori* during gastruloid extension, will necessitate work on several technological fronts, including the construction of the labelled RNAs and detection systems in ES cells, as well as the optical part using gastruloids. Gregor and Duboule laboratories have already started to set up this complex approach in ES cells, which will have to be subsequently transferred into pseudo-embryos. As there is no *a priori* reason for why this should not be transferable to murine cells, we are confident that limited troubleshooting will be sufficient to resolve existing issues.

On the modeling front, Tkacik's group has a proven track record in applying inverse models[111], including maximum entropy and force inference[112,113], although not extensively on polymer data. The polymer model proposed in I.3 is technically feasible due to available community resources such as scalable, open-source simulation codes that we plan to adapt[79,80]. Because the model should handle quantitative data across a range of spatial and temporal scales, it presents a theoretical and technical challenge and thus carries an associated risk. We therefore formulate modeling in this project in a stepwise fashion, which generates insights and predictions along the way right from the start and allows iterative improvements to the model (also via dedicated experiments informed by wrong predictions). This ensures that our modeling approach does not have a single point of failure.

In **Phase II**, experiments involving CRISPR/Cas9-based genome editing in ES cells to produce mutant gastruloids should be straightforward and are well under control in our laboratories. These

include deletions, inversions, or insertions of genetic material through homologous recombination, even though recombination of large DNA fragments remains to be optimized[114]. The analyses of these mutant configurations should be very similar to those under Phase I. However, the production and use of a cohesin complex with light-sensitive rings and of a light sensitive CTCF version will be difficult tasks to realize. They involve the development of novel biology approaches, which can fail at many levels. We nevertheless believe that we can mitigate the risk by first looking at interaction profiles, a straightforward readout, before aiming at simultaneous detection of nascent transcripts. We think that the ERC SyG is an ideal vehicle allowing us to invest the time designing such a tool, which we believe has enormous potential also beyond this grant: side-by-side observation of cells that may or may not have a functional depletion of CTCF or cohesin rings.

The last part of the grant related to the observation and modifications of transcriptional condensates is by and large the most demanding part of this proposal and will require several years of R&D (already started) to both visualize and perturb condensates at select genomic loci. Nevertheless, the potential benefit of such an approach to understand transcription in a more quantitative manner is timely and worth the effort. ERC is likely the only funding opportunity in the EU for collaborative work at the scale required to address such a complex question.

Synergistic elements

The overarching synergistic value of this application is rooted in the overlapping scientific expertise and interests of the three PIs' laboratories. The proposal represents a response to a pressing need, in the field of gene transcription and cell specification, to integrate genomic and imaging data sets in a quantitative manner, leading to model predictions that can be tested by perturbation experiments. The proposal lies exactly at the interface of the three domains of the PIs and connects physics and biology. Its execution requires multiple cutting-edge technologies in genomics, imaging, analysis, and modeling, which manifestly cannot be handled by a single laboratory. We believe that the development and success of such an ambitious interdisciplinary program is only possible within the framework of an ERC SyG. Synergistic effects are supplemented by a close geographical proximity between the two wet experimental poles of the program, which may even lead to the sharing of lab space and facilities on a daily basis. The three PIs have a previous track record of productive pairwise engagements, which the proposed SyG could elevate to a new level, by supporting a joint, large-scale project spanning the entirety of the three groups.

Timeline

Aim	Task	Deliverable	Y1	Y2	Y3	Y4	Y5	Y6
I.1	I.1.1	High resolution 3D profiles of micro-dissected gastruloids and locus selection						
	I.1.2	Modeling dynamic high resolution 3D contact maps for selected loci						
	I.1.3	Dynamic profiling of molecular players involved in chromatin organization						
I.2	I.2.1	Establish gastruloids as a quantitative model for *in vivo* fluorescence imaging						
	I.2.2	Multi-color imaging of CREs and transcription in cell culture and gastruloids						
	I.2.3	Force-inference modeling of physical forces acting on CREs						
I.3	I.3.1	Simulations of DNA polymer dynamics with time-modulated molecular markers						
	I.3.2	Integration of genomic and imaging data to constrain model parameters						
	I.3.3	Identify structural and dynamical motifs controlling transcription						
II.1	II.1.1	CRISPR/Cas9 modifications of CREs in selected loci						
II.2	II.2.1	Development of a photocleavable cohesin complex						
	II.2.2	Development of photo-depletable CTCF for rapid conditional inactivation						
II.3	II.3.1	Site-specific characterization and disassembly of endogenous transcription condensates						

Gregor Duboule Tkacik

References

1. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
2. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
3. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472 (2012).
4. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
5. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
6. Xiao, J. Y., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large changes in transcription. *Elife* **10**, (2021).
7. Ong, C.-T. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
8. Soshnikova, N., Montavon, T., Leleu, M., Galjart, N. & Duboule, D. Functional analysis of CTCF during mammalian limb development. *Dev Cell* **19**, 819–830 (2010).
9. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305--320.e24 (2017).
10. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
11. Chen, H. *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
12. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of Transvection in Living Drosophila Embryos. *Mol. Cell* **70**, 287-296.e6 (2018).
13. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* **76**, 473-484.e7 (2019).
14. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, (2019).
15. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).
16. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
17. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
18. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
19. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **67**, 837-852.e7 (2017).
20. Darbellay, F. & Duboule, D. Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. in *Current Topics in Developmental Biology* **116**, 299–314 (2016).
21. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
22. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–35 (2003).
23. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–5 (2010).
24. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
25. Amândio, A. R. *et al.* Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes Dev.* **35**, 1490–1509 (2021).
26. Turner, D. A. *et al.* Anteroposterior polarity and elongation in the absence of extraembryonic tissues and spatially localised signalling in Gastruloids, mammalian embryonic organoids. *Development* **144**, 3894–3906 (2017).
27. Beccari, L. *et al.* Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).

28. Veenvliet, J. V *et al.* Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* **370**, (2020).
29. van den Brink, S. C. *et al.* Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* (2020). doi:10.1038/s41586-020-2024-3
30. Tiana, G. & Giorgetti, L. Integrating experiment, theory and simulation to determine the structure and dynamics of mammalian chromosomes. *Curr. Opin. Struct. Biol.* **49**, 11–17 (2018).
31. Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci.* **113**, 12168–12173 (2016).
32. Messelink, J. J. B., van Teeseling, M. C. F., Janssen, J., Thanbichler, M. & Broedersz, C. P. Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales. *Nat. Commun.* **12**, 1963 (2021).
33. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
34. Bauer, B. W. *et al.* Cohesin mediates DNA loop extrusion by a “swing and clamp” mechanism. *Cell* **184**, 5448-5464.e22 (2021).
35. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).
36. Amândio, A. R., Lopez-Delisle, L., Bolt, C. C., Mascrez, B. & Duboule, D. A complex regulatory landscape involved in the development of mammalian external genitals. *Elife* **9**, e52962 (2020).
37. Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **181**, 1062-1079.e30 (2020).
38. Schibler, A. *et al.* Histone H3K4 methylation regulates deactivation of the spindle assembly checkpoint through direct binding of Mad2. *Genes Dev.* **30**, 1187–1197 (2016).
39. Dickel, D. E. *et al.* Function-based identification of mammalian enhancers using site-specific integration. *Nat. Methods* **11**, 566–571 (2014).
40. Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* **13**, 1009–1011 (2016).
41. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539--553.e8 (2020).
42. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
43. Imakaev, M. V., Fudenberg, G. & Mirny, L. A. Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.* **589**, 3031–3036 (2015).
44. Zhang, B. & Wolynes, P. G. Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci.* **112**, 6062–6067 (2015).
45. Fiorillo, L. *et al.* Inference of chromosome 3D structures from GAM data by a physics computational approach. *Methods* **181–182**, 70–79 (2020).
46. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
47. Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* **116**, 683–698 (2004).
48. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–8 (2011).
49. Sato, H., Das, S., Singer, R. H. & Vera, M. Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annu. Rev. Biochem.* **89**, 159–187 (2020).
50. Saad, H. *et al.* DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living Cells. *PLoS Genet.* **10**, (2014).
51. Germier, T. *et al.* Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys. J.* **113**, (2017).
52. Xu, H., Sepúlveda, L. A., Figard, L., Sokac, A. M. & Golding, I. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nat. Methods* **12**, 739–742

(2015).
53. Little, S. C., Tikhonov, M. & Gregor, T. Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. *Cell* **154**, 789–800 (2013).
54. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
55. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* (2016).
56. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* **74**, 212-222.e5 (2019).
57. Dubarry, N., Pasta, F. & Lane, D. ParABS systems of the four replicons of Burkholderia cenocepacia: New chromosome centromeres confer partition specificity. *J. Bacteriol.* **188**, 1489–1496 (2006).
58. Bystricky, K. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. *FEBS Lett.* **589**, 3014–3022 (2015).
59. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
60. Liu, Z., Lavis, L. D. & Betzig, E. Imaging live-cell dynamics and structure at the single-molecule level. *Mol Cell* **58**, 644–659 (2015).
61. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
62. Dunsby, C. Optically sectioned imaging by oblique plane microscopy. *Opt. Express* **16**, 20306 (2008).
63. Botcherby, E. J., Juškaitis, R., Booth, M. J. & Wilson, T. An optical technique for remote focusing in microscopy. *Opt. Commun.* (2008). doi:10.1016/j.optcom.2007.10.007
64. Bouchard, M. B. *et al.* Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms. *Nat. Photonics* **9**, 113–119 (2015).
65. Voleti, V. *et al.* Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nat. Methods* **16**, 1054–1062 (2019).
66. Yang, B. *et al.* Epi-illumination SPIM for volumetric imaging with high spatial-temporal resolution. *Nat. Methods* **16**, 501–504 (2019).
67. Sapoznik, E. *et al.* A versatile oblique plane microscope for large-scale and high-resolution imaging of subcellular dynamics. *Elife* **9**, 1–39 (2020).
68. Gregor, T., Garcia, H. G. & Little, S. C. The embryo as a laboratory: quantifying transcription in Drosophila. *Trends Genet.* **30**, 364–375 (2014).
69. Zhou, H. Y. *et al.* A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* **28**, 2699–2711 (2014).
70. Zuin, J. *et al.* Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv* 2021.04.22.440891 (2021). doi:10.1101/2021.04.22.440891
71. El Beheiry, M., Dahan, M. & Masson, J.-B. InferenceMAP: mapping of single-molecule dynamics with Bayesian inference. *Nat. Methods* **12**, 594–595 (2015).
72. Pérez García, L., Donlucas Pérez, J., Volpe, G., V Arzola, A. & Volpe, G. High-performance reconstruction of microscopic force fields from Brownian trajectories. *Nat. Commun.* **9**, 5166 (2018).
73. Ferretti, F., Chardès, V., Mora, T., Walczak, A. M. & Giardina, I. Building General Langevin Models from Discrete Datasets. *Phys. Rev. X* **10**, 31018 (2020).
74. Grah, R., Zoller, B. & Tkačik, G. Nonequilibrium models of optimal enhancer function. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31614–31622 (2020).
75. Zoller, B., Little, S. C. & Gregor, T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell* **175**, 835-847.e25 (2018).
76. Polovnikov, K. E., Gherardi, M., Cosentino-Lagomarsino, M. & Tamm, M. V. Fractal Folding and Medium Viscoelasticity Contribute Jointly to Chromosome Dynamics. *Phys. Rev. Lett.* **120**, 88101 (2018).
77. Lampo, T. J., Kennard, A. S. & Spakowitz, A. J. Physical Modeling of Dynamic Coupling between Chromosomal Loci. *Biophys. J.* **110**, 338–347 (2016).
78. Frisch, M. J., Pople, J. A. & Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **80**, 3265–3269 (1984).

79. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
80. Anderson, J. A., Glaser, J. & Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* **173**, 109363 (2020).
81. Brackley, C. A., Marenduzzo, D. & Gilbert, N. Mechanistic modeling of chromatin folding to understand function. *Nat. Methods* **17**, 767–775 (2020).
82. Bottou, L. Stochastic Gradient Descent Tricks. in *Neural Networks: Tricks of the Trade* (eds. Montavon, G. et al.) **7700**, 421–436 (2012).
83. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by Simulated Annealing. *Science (80-.).* **220**, 671–680 (1983).
84. Brackley, C. A. *et al.* Complex small-world regulatory networks emerge from the 3D organisation of the human genome. *Nat. Commun.* **12**, 5756 (2021).
85. Bianco, S. *et al.* Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, 662–667 (2018).
86. Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–3 (2003).
87. Darbellay, F. *et al.* The constrained architecture of mammalian Hox gene clusters. *Proc. Natl. Acad. Sci.* **116**, 13424–13433 (2019).
88. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb. Perspect. Biol.* a040147 (2021). doi:10.1101/cshperspect.a040147
89. Xiang, J.-F. & Corces, V. G. Regulation of 3D chromatin organization by CTCF. *Curr. Opin. Genet. Dev.* **67**, 33–40 (2021).
90. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
91. Pauli, A. *et al.* Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev. Cell* **14**, 239–251 (2008).
92. Zhang, W. *et al.* Optogenetic control with a photocleavable protein, PhoCl. *Nat. Methods* **14**, 391–394 (2017).
93. Lee, D. *et al.* Temporally precise labeling and control of neuromodulatory circuits in the mammalian brain. *Nat. Methods* **14**, 495–503 (2017).
94. Wang, H. & Hahn, K. M. LOVTRAP: A Versatile Method to Control Protein Function with Light. *Curr. Protoc. cell Biol.* **73**, 21.10.1-21.10.14 (2016).
95. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–21 (2015).
96. Anania, C. *et al.* In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *bioRxiv* 2021.04.14.439779 (2021). doi:10.1101/2021.04.14.439779
97. Niopek, D., Wehler, P., Roensch, J., Eils, R. & Di Ventura, B. Optogenetic control of nuclear protein export. *Nat. Commun.* **7**, 10624 (2016).
98. Chen, S. Y. *et al.* Optogenetic Control Reveals Differential Promoter Interpretation of Transcription Factor Nuclear Translocation Dynamics. *Cell Syst.* **11**, 336-353.e24 (2020).
99. Kögler, A. C. *et al.* Extremely rapid and reversible optogenetic perturbation of nuclear proteins in living embryos. *Dev. Cell* **56**, 2348-2363.e8 (2021).
100. Singh, A. P. *et al.* Optogenetic control of the Bicoid morphogen reveals fast and slow modes of gap gene regulation. *bioRxiv* 2021.10.13.464280 (2021). doi:10.1101/2021.10.13.464280
101. de Mena, L., Rizk, P. & Rincon-Limas, D. E. Bringing Light to Transcription: The Optogenetics Repertoire. *Front. Genet.* **0**, 518 (2018).
102. Vazquez, J., Belmont, A. S. & Sedat, J. W. Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr. Biol.* **11**, 1227–1239 (2001).
103. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
104. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, 387–392 (2018).
105. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity

enhancers. *Elife* **6**, e28975 (2017).
106. Chen, H. *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
107. Barinov, L., Ryabichko, S., Bialek, W. & Gregor, T. Transcription-dependent spatial organization of a gene locus. *arXiv* **2012.15819**, (2020).
108. Gräf, R. Maltose-Binding Protein as a Fusion Tag for the Localization and Purification of Cloned Proteins in Dictyostelium. *Anal. Biochem.* **289**, 297–300 (2001).
109. Reed, E. H., Schuster, B. S., Good, M. C. & Hammer, D. A. SPLIT: Stable Protein Coacervation Using a Light Induced Transition. *ACS Synth. Biol.* **9**, 500–507 (2020).
110. Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175**, 1481-1491.e13 (2018).
111. Tkačik, G. & Bialek, W. *Information Processing in Living Systems*. *Annual Review of Condensed Matter Physics* **7**, (2014).
112. Tkačik, G. *et al.* Searching for collective behavior in a large network of sensory neurons. *PLoS Comput. Biol.* **10**, e1003408 (2014).
113. Brückner, D. B., Ronceray, P. & Broedersz, C. P. Inferring the Dynamics of Underdamped Stochastic Systems. *Phys. Rev. Lett.* **125**, 58103 (2020).
114. Ioannidi, E. I. *et al.* Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. *bioRxiv* 2021.11.01.466786 (2021). doi:10.1101/2021.11.01.466786