**DATA ANALYSIS AND MODELING**

**Title:** **Multi-scale Modeling and Functional Annotation of 3D genome organization**

**PROJECT SUMMARY**

Comprehensive understanding of three-dimensional (3D) genome organization requires computational advances in simulations, structural ensemble comparison, and functional genomic analysis. ChIA-PET experiments, together with other data types, provide multiscale information that can resolve these issues. ChIA-PET data clusters, for example, reveal localized protein-mediated structures, while ChIA-PET singleton data reveal longer-range effects from structures occurring weakly in the ensemble. This diversity of hierarchical behaviors underscores the value of multiscale approaches in 3D genomic data analysis. We recently implemented a multidimensional scaling (MDS) algorithm to rapidly and robustly determine structures from ChIA-PET data, and we have also built a web-based computational platform called the 3D Nucleome Modeling Engine **(3D-NOME)** to model and visualize structures at multiple scales. In this Data Analysis and Modeling proposal, we will seek to further advance our understanding of 3D genome organization by 1) Optimizing the 3D-NOME modeling platform to rapidly calculate 3D structures at multiple scales from combined ChIA-PET and nuclear imaging data using advanced MDS techniques; 2) Developing computational methods to compare and evaluate structural ensembles across different conditions, as well as biophysically modeling the dynamics of structural ensembles at the local and chromosome-wide levels; and 3) Annotating 3D structures with functional genomics datasets to discover spatial functional compartments, and integrating DNA-DNA interaction networks with 3D structures to study their functional relevance and significance at the systems level. These tools and resources will enable a wide array of researchers to revisit their functional systems and discoveries under the light of the spatiotemporal 4D nucleome.

**SPECIFIC AIMS**

Current advances in next-generating sequencing techniques (e.g., ChIA-PET and RICH-PET) and nuclear imaging methods provide an unprecedented opportunity to decode the human nucleome—by modeling its spatial structures at various levels spanning a hierarchical three-dimensional (3D) structure (whole chromosome, segment, and particular loci), by studying the dynamics of its spatial structures across conditions and across time, and by integrating its structures with functional genomics datasets for functional interpretation. Such decoding of the nucleome raises unique computational challenges that require advances in data pre-processing, structural data modeling, data visualization, and data integration. To address these unique computational needs and challenges, we will build an integrative and comprehensive data-analysis platform we call the 3D Nucleome Modeling Engine (**3D-NOME**). With this computational engine, first, we will refine the hierarchical 3D structural nucleome models by integrating ChIA-PET datasets with nuclear imaging datasets and by implementing advanced multi-dimensional scaling (MDS) methods. Second, we will develop novel strategies to enable comparisons between 3D structures derived across different time points and conditions during cell differentiation. This will enable us to compare and contrast nucleome structures and their dynamics under different conditions, such as across time points during cellular differentiation. Third, we will integrate inferred nucleome structures with functional genomics datasets from public repositories including ENCODE and Roadmap Epigenomics projects as well as from experiments that we will conduct in human immune cells as part of this program. This integration will enable us to study and understand the functional meaning of 4D nucleome structures at the individual component-level as well as at the systems level.

**Aim 1. Engineer the multi-scale modeling platform for 3D genome structure prediction and visualization:** We will develop our open-source, multi-scale modeling platform to provide accurate 3D models of genome structure, incorporating novel experimental data at multiple spatial scales. First, we will optimize the simulation and sampling engine to handle nuclear imaging data from 3D-FISH microscopy at the chromosomal level. Second, we will develop more rapid computational methods to calculate structures from this data using multidimensional scaling techniques. Third, we will integrate these approaches with those we developed for ChIA-PET and other data to optimize our methods and refine our inferred structures.

**Aim 2. Develop computational methods for comparing structure ensembles**: Understanding the functional importance of the 4D nucleome requires improvement in the visualization and analysis of structural ensembles. To address this issue, we will first develop methods for interactive analysis of a structural ensemble associated with a specific experimental condition. Second, we will develop methods for comparison of ensembles across conditions, including identification of critical substructures robust across conditions. We will analyze these critical substructures in detail by dynamic modeling using Langevin dynamics accounting for polymer entanglement. Finally, we will investigate the dynamics of the structures macroscopically using rule-based modeling approaches that are efficient at large scales.

**Aim 3. Functionally annotate the 3D nucleome structure:** To interpret the functional relevance and significance of the 4D nucleome structures, we will first annotate the inferred spatial structures with functional genomics datasets, such as transcriptomic, epigenomic datasets and GWAS SNPs. Second, we will incorporate the 3D information into the predictive machine-learning models we have built to understand regulatory mechanisms. Third, we will integrate DNA-DNA interaction networks obtained from ChIA-PET datasets with the spatial structures, to reveal the system-wide impact of functional spatial structures.

Successful completion of this data analysis and modeling component will not only provide novel computational methods to model a 4D nucleome at multiple scales by integrating ChIA-PET and nuclear imaging datasets, but will also enable studying the dynamics and functional relevance of the inferred spatial structures. In addition, we will also develop and disseminate a comprehensive computational tool, 3D-NOME, to study, visualize, and interpret the 4D nucleome structures across diverse conditions and at multiple spatial scales. This tool will enable a wide array of researchers to revisit their functional systems and discoveries under the light of the spatiotemporal 4D nucleome.

**A video demonstrating the 3D models and the functionality of the visualization tool in the 3D-NOME platform is available as supplemental material to the application.**

**RESEARCH STRATEGY**

**SIGNIFICANCE**

Advances in mathematical and computational formalisms are necessary to address fundamental issues in 4D nucleome analysis. Issues include the multi-scale properties of populations of cells, the internal structure of cells ([1](#_ENREF_1), [2](#_ENREF_2)) and the cell nucleus ([3](#_ENREF_3), [4](#_ENREF_4)), the dynamic understanding of structural ensembles, and the integration of 3D structural information with functional genomic data. More efficient new methods are particularly needed for biophysical modeling of chromatin from improved experimental data such as multi-scale ChIA-PET and nuclear imaging. Although prior methods for estimating 3D structure have been developed, many are computationally expensive ([5-7](#_ENREF_5)). One promising approach to alleviate this issue is multidimensional scaling (MDS) ([8](#_ENREF_8)), a computational technique to transform pairwise similarity scores among objects into Euclidean locations for each object. MDS was recently adapted to Hi-C data to rapidly generate coarse models of nucleome structure ([9](#_ENREF_9)). Nevertheless, establishing not only the accuracy and robustness of such methods but also their relevance for functional prediction remain key challenges. Developing analogous fast methods for ChIA-PET and nuclear imaging data would be of great value to validate and improve structure inference. Beyond predictions of individual structures and structural ensembles, using these predictions to identify functionally important substructures is another open challenge. This is a crucial problem analogous to determining active sites in proteins ([10](#_ENREF_10), [11](#_ENREF_11)). Finally, once genomic structures and ensembles are understood, they can be integrated into broader approaches for functional genomic interpretation, e.g., through application of predictive models and integrative interaction network models. These types of computational advances are critical to translating 3D structural data to enable understanding of cellular functions.

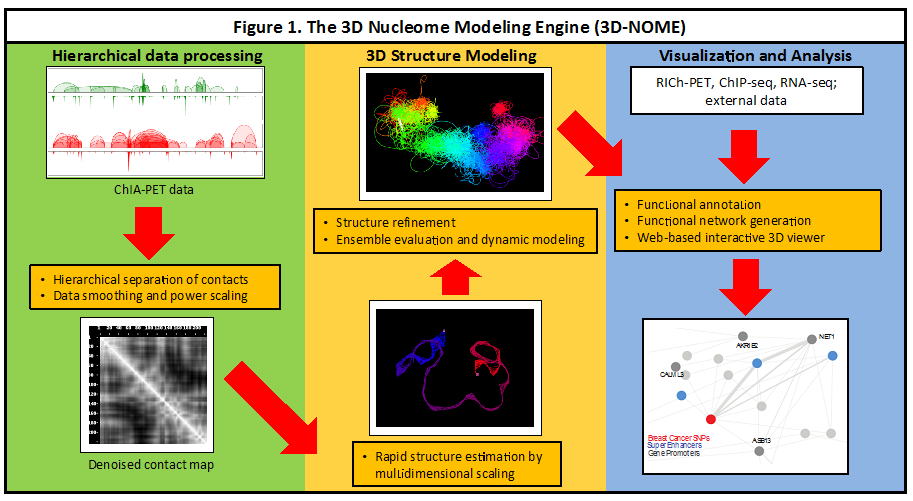


Fig 1. Schematic of our 3 Dimensional NucleOme Modeling Engine (3D-NOME) platform. We have implemented methods for hierarchical data pre-processing, structure modeling by multidimensional scaling, and visualization based on ChIA-PET data. We propose to improve 3D-NOME by algorithmic optimizations, adaptation for new data types, novel ensemble structure analysis, and functional annotation of spatial structures. A video showing the functionality of the 3D visualization tool in the 3D-NOME platform is available as a supplemental material to the application.

**INNOVATION**

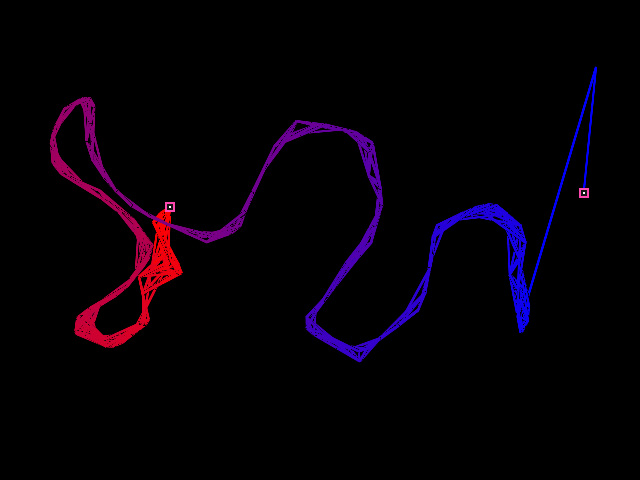
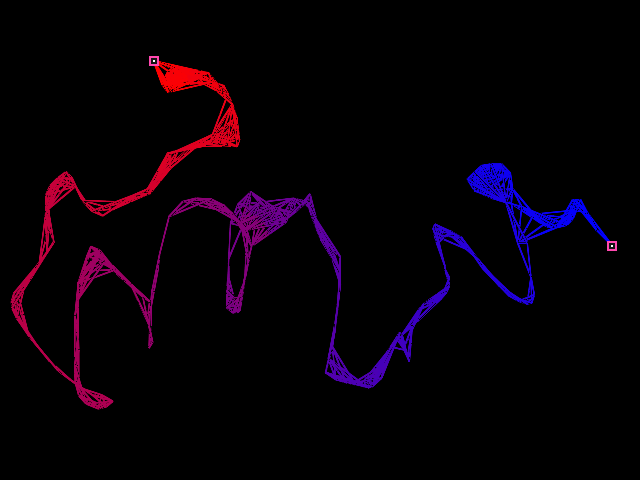
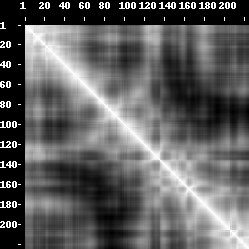
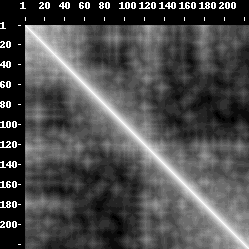
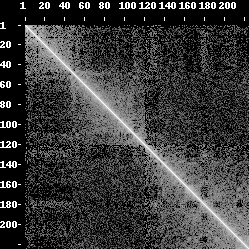
Our proposal contains multiple innovative aspects to advance understanding of, and technological approaches for, inference and interpretation of the 4D nucleome. ***1)*** We will further develop and optimize a modeling platform—**3-Dimensional NucleOme Modeling Engine (3D-NOME)**—that we have built in order to calculate and visualize genome structures **(Fig. 1).** Key capabilities that we have incorporated into 3D-NOME are novel approaches for hierarchical denoising and novel MDS algorithms that greatly increase the speed at which structures can be generated from ChIA-PET data. These are novel approaches that have been developed for ChIA-PET data. We will further innovate by providing the first hierarchical de-noising and MDS approaches for nuclear imaging data. By comparing results from different data types, we will optimize the MDS approach. Our methods will be developed as open-source software to provide a multi-scale computational and web-based visualization platform that can render and compare genomic structures, addressing a pressing usability and accessibility problem in 3D genome analysis. ***2)*** We will develop novel computational tools to evaluate robustness within and between structural ensembles, bridging the gap between structure inference and functional interpretation. These ensemble analyses will be combined with novel dynamic modeling approaches to investigate the impacts of polymer entanglement and protein diffusion, and will be augmented by rule-based approaches to study dynamics at chromosomal scales. ***3)*** Finally, we will develop computational approaches to annotate 3D structures with functional genomics datasets to reveal the functional relevance of these spatial structures.

**APPROACH**

***Preliminary Results***. We have developed a comprehensive software platform, 3D-NOME, for simulation and visualization (**Fig. 1**). 3D-NOME includes three components: 1) hierarchical denoising, 2) multidimensional scaling and refinement for rapid structure inference, and 3) web-based visualization tools. Our pipeline provides significant advances over the current state-of-the-art in each of these aspects, as described below.

*Hierarchical denoising and transformation of contact maps.* While the contacts from ChIA-PET or HiC data provide information for reconstructing structures, noise in contact map data can also yield inconsistencies in inferred structures. We have developed a procedure to denoise ChIA-PET contact data, with the innovation that we perform this hierarchically, i.e., denoise iteratively at large (~10Mb) down to short (~kb) length scales. Our method improves upon that of Lesner et al ([9](#_ENREF_9)), in which denoising is done before MDS but without hierarchical separation. In brief, we transform contact frequencies between nodes into 3D distances based on scaling laws from polymer physics ([12](#_ENREF_12" \o "Varoquaux, 2014 #14)), smooth these 3D distance values, then revert the 3D distances back into contact frequencies. Further denoising is performed by power scaling of contact frequencies, where the power scaling exponent and a scale factor are key parameters **(Fig. 2).** The scaled contact frequencies are used to generate a connectivity graph, and then graph distances are used as input to the MDS method described below. A further novelty of our approach is that we generalize to fuzzy graphs ([13](#_ENREF_13)), which we found provide a more effective graph distance function.

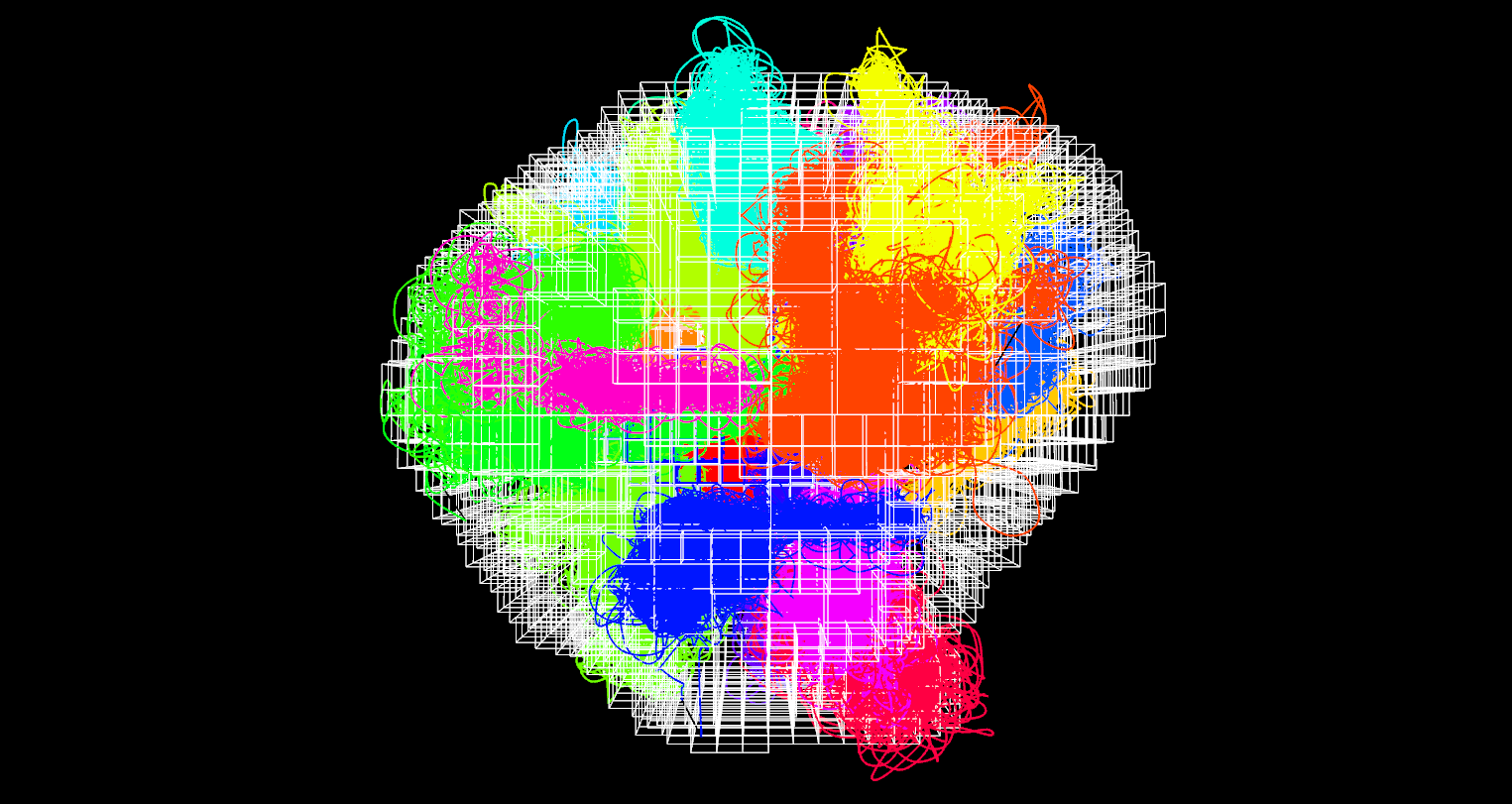
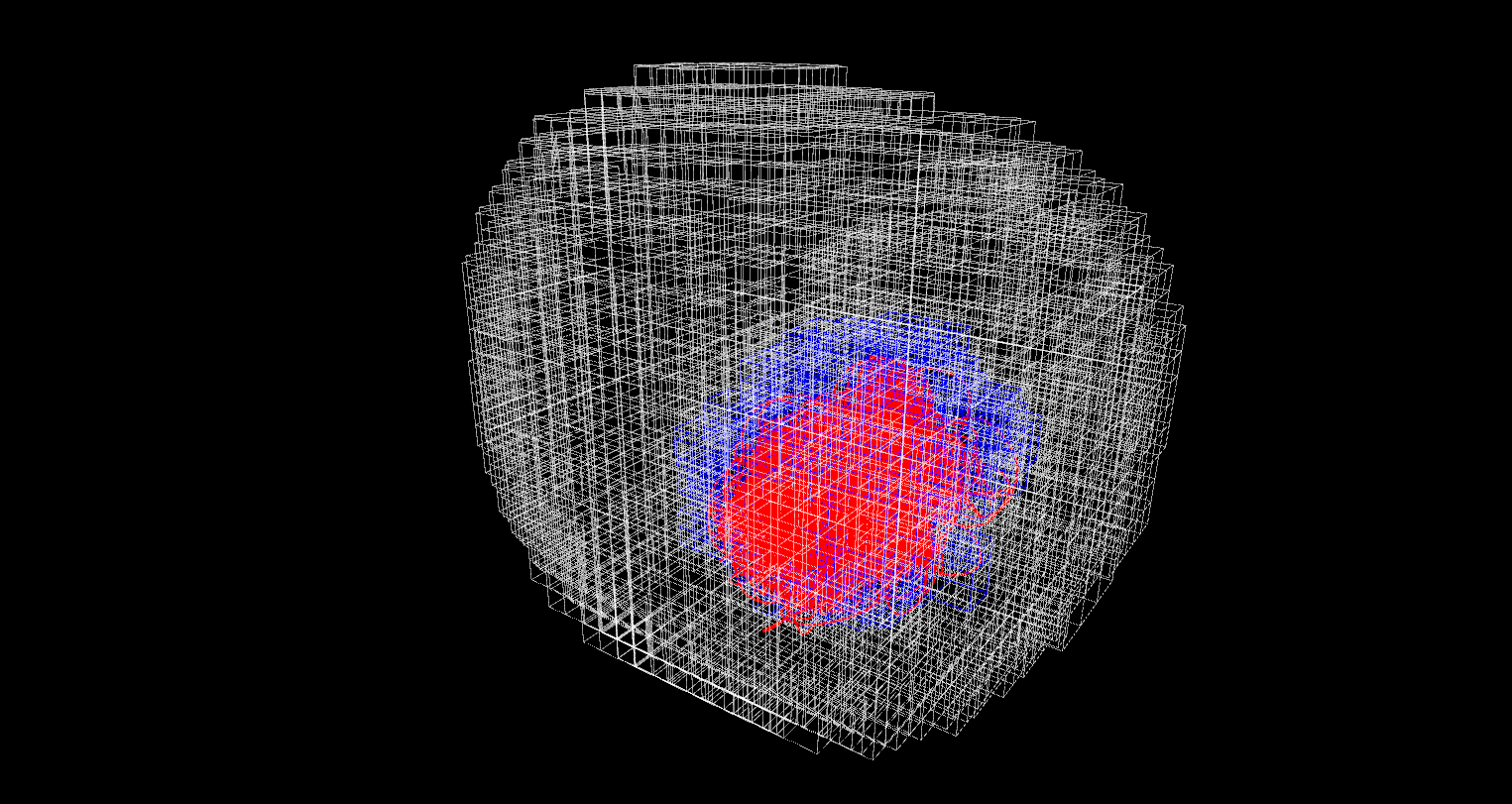
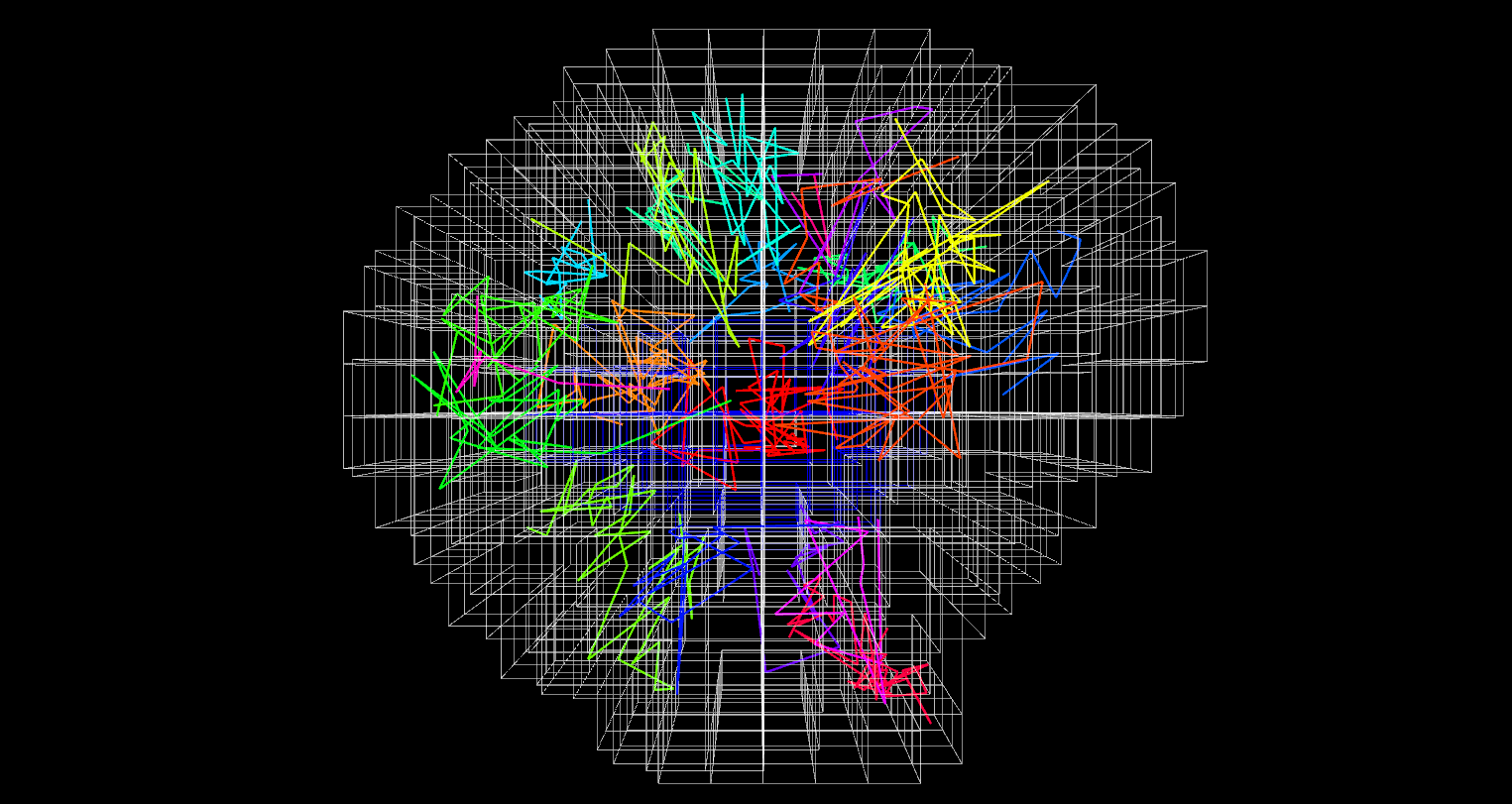
*Multidimensional scaling (MDS) using ChIA-PET data provides rapid and improved structure prediction.*MDS greatly speeds up estimates of chromosome conformations. We developed the first MDS algorithm for ChIA-PET data that is run on the graph distances from the hierarchical denoising above. MDS minimizes the error between a similarity (distance) matrix and the matrix that would be observed given Euclidean locations of all nodes. At each step, the algorithm moves N points in 3D and exits if no further improvement is made. Because of the large number of simultaneous moves, MDS is able to estimate coarse structure much more rapidly than prior methods based on single-move Monte Carlo approaches ([9](#_ENREF_9)). In our preliminary results, we found that we can achieve structures for ChIA-PET data with average 20sec calculation time on a stand-alone PC workstation. In addition, we performed extensive parameter testing (e.g., for the exponent k in the power scaling of contact frequencies) and found that structure estimates are sensitive to these parameters **(Fig. 2).** Values of the exponent parameter from HiC yield poor results for ChIA-PET data. Moreover when compared to Hi-C data, the contact maps from ChIA-PET data are characterized by lower noise. For example, distances between neighboring nodes along the chromosome should be equidistant. However, the variation in these distances is 11% higher in HiC inferred structures than in ChIA-PET inferred structures, indicating greater uncertainty in spatial positioning from HiC data.



**Figure 2.** (Top) Inferred structures are sensitive to denoising parameters. (Left) Chr1 with parameters: scaling factor = 0.65, exponent = 95.0 (Right) Chr8 with parameters: scaling factor = 0.55, exponent = 115.0. (Bottom). As the exponent parameter increases, contact frequencies becoming increasingly binary, providing a denoising effect. (Left) Chromosome 1 contact map drawn for raw ChIA-PET data, (Center) Contact map after power scaling. (Right) Contact map after MDS procedure. Note that the 2D map becomes less random and more similar to a polymer type. 2D maps drawn for chr1 with scaling factor = 1.0, exponent = 35.0.

*Refinement of MDS structures.* Because MDS yields a coarse structure estimate, we implemented a restraint-based 3D-modeling approach ([12](#_ENREF_12)) that refines the MDS results by single-move Monte Carlo approaches. This refinement is performed by transforming ChIA-PET interaction frequencies into preferred spatial distances, and then minimizing an energy function based on the deviation from the preferred spatial distances. We have also implemented a more computationally intensive Langevin dynamics ([14](#_ENREF_14)) approach, which can yield a full ensemble of structures rather than a single optimized structure, as described in Aims 2.3 and 2.4.

*Visualization and software engineering.* We developed a web-based GPU accelerated viewer to allow visualization of predicted structures in any browser. The viewer is fully interactive, with mouse-controlled translations, rotations, and zooms. The viewer provides a range of options to aid in structural analysis, such as the ability to view the whole nucleus and each chromosome at adjustable resolution, to highlight regions of interest based on genomic coordinates, and to indicate the binding locations of protein factors used to generate the ChIA-PET library. These features have been implemented using WebGL technology and tested on stand-alone versions implemented in C/C++/QT, java and python bindings. The website provides the option of displaying multiple chromosomes in side-by-side viewers for convenient visual comparison of the same genomic region in different cell lines, or under different treatment conditions. The software architecture for 3D-NOME has been designed by software engineers at JAX and is divided into components for Modeling and Optimization (denoising and structure inference); Web Services (data input/output and storage); and Web Application (3D visualization). A video of visualization tool is in supplement.



***Strategy***

**Aim 1. Engineer the multi-scale modeling platform for 3D genome structure prediction and visualization**

We propose to develop methods that model chromatin structure from nuclear bioimaging data and then implement a multidimensional scaling (MDS) approach for this prediction. We will then compare the results of ChIA-PET and nuclear bioimaging data to validate chromatin structures and refine these methods.

*1.1. Modeling chromosome structure from nuclear imaging data.* 3D-FISH measurements provide valuable data for improved structure determination, and we propose to implement constraints from 3D-FISH for this purpose. In this technique, designed fluorescent probes mark the arms of chromosomes. We will develop and optimize an algorithm to perform image segmentation, i.e., estimation of surfaces of the regions containing each chromosome and the overall nucleus. This surface will be translated to a volume and denoised for image segmentation artifacts. To represent volume we will use an octree, a tree-based data structure commonly used to partition 3D space. We have already developed a prototype simulation with two constraining volumes: one for a painted chromosome and one for the cell nucleus. Points on the chromosome are constrained to the specified volume and other chromosomes are prohibited. Nuclear constraint restricts the movement of all genomic loci. We have implemented an initial single-move Monte Carlo (non-MDS) simulation based on 3D-FISH data from a GM12878 cell line with spatial constraints enforced on chromosome 1 and the nucleus (**Fig. 3**). Future data will involve simultaneous 3D-FISH from multiple labeled chromosomes, and we propose to generalize our image segmentation and structure inference for these constraints.

Figure 3. Chromosomal structures inferred by 3D-NOME under constraint by both the chromosome painting surfaces and whole nucleus border. Octrees representing the constrained volume are shown in white (nucleus) and blue (chromosome 1) grids. (Top) The whole nucleus model; (middle) subchromosomal level (bottom) subanchor levels.

*1.2. Multidimensional scaling for nuclear imaging data.*Currently we use the slow method of single-move Monte Carlo to infer structures from 3D-FISH bioimaging. Here we propose to adapt the hierarchical denoising and MDS steps for this type of data to yield fast estimates of structures. This process can be performed by initially by MDS in the absence of constraining volumes, then using techniques similar to simulated annealing to progressively constrain the volume. The chromosomal volume diameter is a key binding factor in this protocol, as it provides a scaling constant for the MDS algorithm. We will also expand our approach to make use of chromatin density data from the 3D-FISH experiments. This problem is analogous to those associated with fitting a protein 3D structure to an electron density map (e.g., as implemented in MinkoFit3D, a method which uses Minkowski sums in 3D protein modeling ([15](#_ENREF_15))). Density data will be analyzed to pinpoint chromatin fragments that are in agreement with global chromatin shape, but do not match the experimentally determined density.

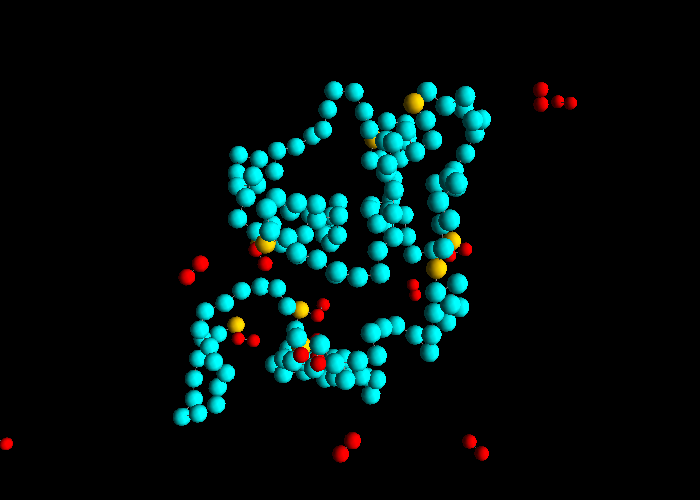
*1.3. Optimization of MDS approaches.*We found that MDS-based structural inferences can be sensitive to parameters such as the exponent *k* in the power scaling for hierarchical denoising, e.g., by altering the exponent parameter, one can change the overall genomic density from densely packed to unfolded. We will examine the impact of this and other parameters on structure determination. We will use commonalities in inferred structural properties among ChIA-PET, 3D-FISH, and Hi-C data to assess best choices of parameter values and robust structural features.

**Aim 2. Computational Methods for Comparing Structure Ensembles**

*2.1. Interactive analysis of structural ensembles.* We will adapt our method for visualization of single structures to visualize ensembles (representative structures, 3D density plots), based on structure alignment performed *via* singular value decomposition using the Kabsch algorithm ([16](#_ENREF_16)). A key software engineering goal will be the development of interactivity between the visualizations and the contact data. We will develop a graphical interface that allows users to select contacts that can be removed or added and then reinitiate the structural ensemble computation, so that the structural impact of individual ChIA-PET or HiC contacts can be visualized.

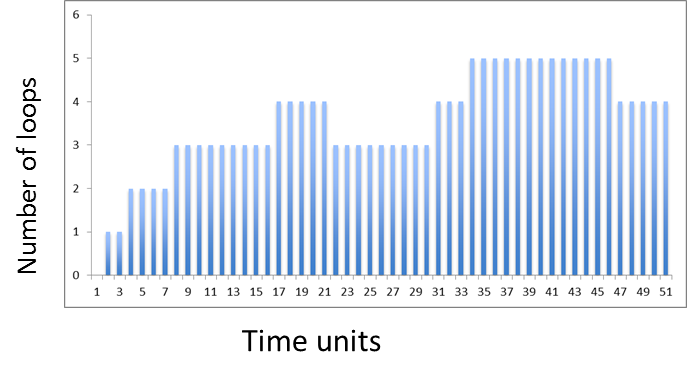
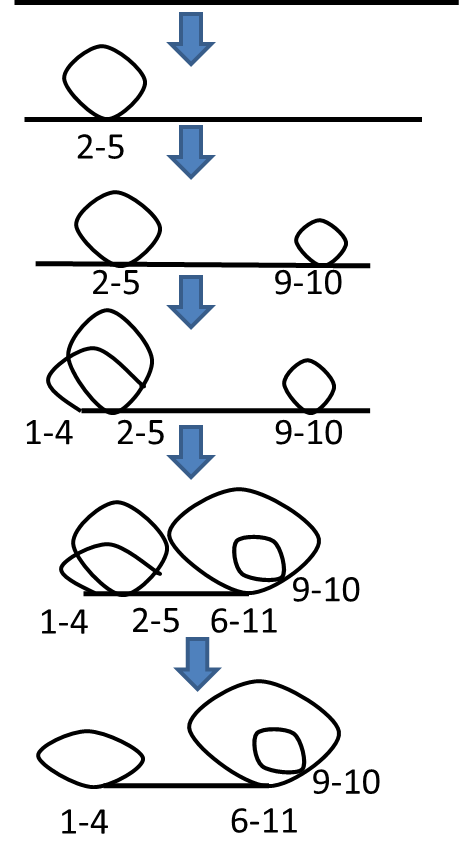
*2.2. Comparison of ensembles.*Comparison of ensembles from different conditions will be critical to biological interpretation. However, alignment of structures from two ensembles will likely be meaningful for only a small subset of regions robustly preserved across conditions. We will develop approaches to resolve this problem by first identifying robust substructures within an ensemble. Based on the MDS approaches of Aim 1, we expect to be able to sample thousands of structures from each ensemble. Pairwise distance relationships and their noise will be computed from these sampled structures, and points with strong fixed relationships will be identified and then merged to define critical substructures within each ensemble. We will normalize for contiguity along the chain based on expected self-avoiding walk and fractal globule behaviors and also assess how much information is gained by structural modeling in comparison to only using measured contact data (ChIA-PET, HiC). We will then develop methods to use critical substructures as representative sets to align within the Kabsch algorithm framework to compare ensembles.

*2.3. Dynamic modeling of critical substructures.* An alternative to the optimization approach of Aim 1 is dynamic modeling, which can more realistically sample from the ensemble. We have implemented an approach that uses Langevin dynamics ([14](#_ENREF_14)) to model the chromosome as hard-sphere excluded-volume balls and springs in the high viscosity limit. This model allows for simulation of diffusing proteins ([17](#_ENREF_17)) **(Fig. 4)** and kinetic mechanisms of insulation and facilitation ([18](#_ENREF_18)). Our method also includes modeling of biochemical reactions, distinguishing it from prior molecular dynamics platforms for chromatin modeling, e.g., OpenMM ([19](#_ENREF_19)). Dynamic modeling is also necessary for assessing polymer entanglement effects, which are responsible for the fractal globule behavior of human chromosome structures ([5](#_ENREF_5)) and which we showed have dramatic effects on structural ensembles of idealized polymers ([20](#_ENREF_20), [21](#_ENREF_21)). Using adaptive time-step polymer simulation models that we have developed to enforce entanglement constraints ([21](#_ENREF_21)), we will simulate the dynamics within critical substructures to determine whether there are synergies/conflicts among the contacts that distinguish experimental conditions.

**Figure 4:** (a) Langevin dynamics simulation workflow. (b) A snapshot from a simulation with a single chromosome. Sites colored orange are binding sites for proteins (red). The DNA binding proteins can bind two chromosomal sites simultaneously, and thereby mediate looping interactions between distant genomic sites.

*2.4. Rule-based modeling of ensembles.* A limitation of the Langevin approach is that it is computationally expensive, making it unsuitable for parameter scans, rapid hypothesis testing, and very long time scales. As a complementary approach, we propose non-spatial rule-based modeling ([22](#_ENREF_22), [23](#_ENREF_23)) to simulate the dynamics of chromatin interactions at larger scales. Rule-based modeling defines biochemical reaction networks by a concise set of rules describing system transitions, and allows for the simulation of systems where combinatorial complexity prevents traditional network construction ([24](#_ENREF_24)). We propose to convert normalized interaction frequencies into relative binding affinities between identified ChIA-PET clusters, construct biochemical reaction networks assuming independent binding sites, and then generate ensembles by running millions of stochastic simulations **(Fig. 5).** The predicted interactions will be used for initial conditions for Langevin dynamics simulations. Conversely, the Langevin dynamics simulations will inform rule-based modeling by identifying prohibited interactions. We will use this framework to model transcriptional activity by linking enhancer-promoter interactions to gene activation, and compare these predictions to RNA-seq data as an independent validation of our structural modeling. Discrepancies will indicate necessary structural refinements.



**Figure 5:** (Left) A single trajectory of loop formation in chromosome 6, based on our rule-based network model. (Right) The number of looping interactions in the system as a function of time (arbitrary time units).

**Aim 3. Functional annotation of the 3D nucleome structure**

In addition to ChIA-PET datasets, we will generate a diverse array of functional genomics datasets, including RICh-PET data to capture RNA-chromatin interactions, RNA-seq data to assess expression patterns and histone modifications, and ATAC-seq data to characterize the epigenetic landscape. In addition to these datasets, we will also make use of the functional genomics datasets generated by large consortia such as ENCODE ([25](#_ENREF_25)), Roadmap Epigenomics ([26](#_ENREF_26)) and Blueprint Epigenome ([27](#_ENREF_27)) for cell types of interest, i.e., GM12878, embryonic stem, T and B cells. To functionally annotate and interpret 3D nucleome structures, we will first annotate the 3D genome with functional genomics datasets to identify functionally relevant spatial compartments. Next, we will incorporate spatial information into our predictive models for understanding regulatory mechanisms. Finally, we will integrate the 3D structures with DNA-DNA interaction networks to identify functionally critical system-level structures and to study the system-wide impact of these structures.

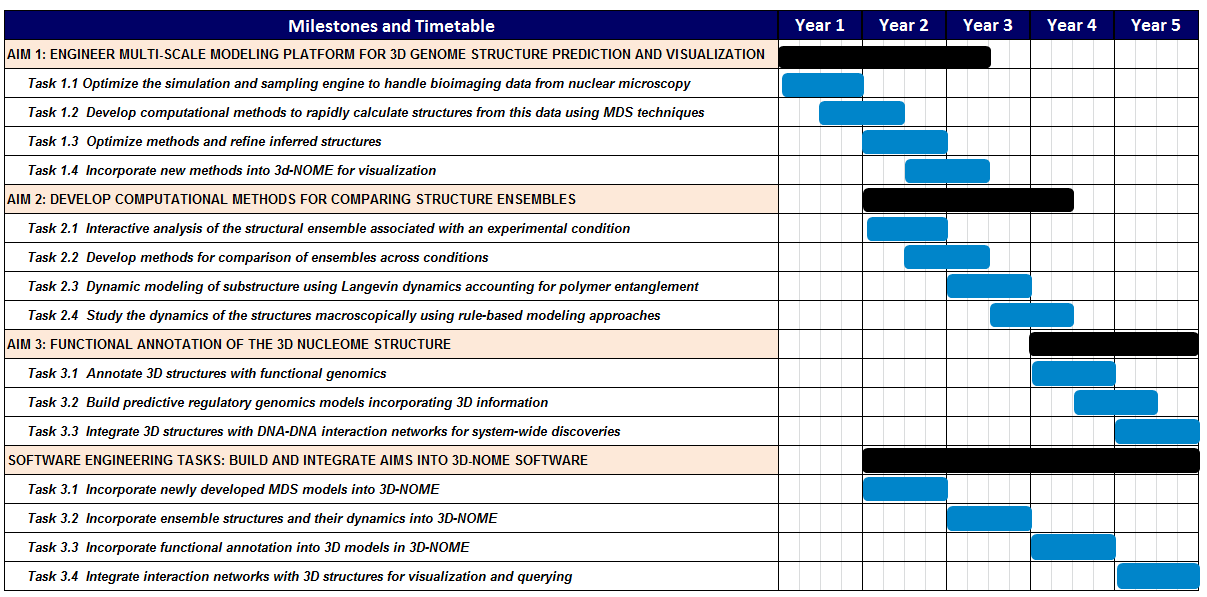
*3.1. Functional annotation and interpretation of the 3D genome.* Interpreting the functional relevance and significance of the 3D nucleome requires integrating 3D structures with diverse transcriptomic and epigenomic datasets, which were traditionally presented in linear (1D) space. We have extensive experience in integrating diverse 1D functional genomics datasets, including TF binding, gene expression and epigenetic datasets to study regulatory mechanisms ([28-32](#_ENREF_28)). We will extend our earlier integrative models for functional genomics by incorporating the information derived from the inferred 3D structures. First, in each cell type, we will annotate the inferred substructures using the genomic coordinates from all available public genomic datasets as well as newly generated datasets as proposed in this project, in order to identify and study functional compartments in the 3D space, e.g., DNA regions that are transcriptionally active/inactive, epigenetically modified and bound by TFs. Such annotation of the 3D genome will enable us to visualize and study functional compartments in the 3D space, as well as to identify genomic and epigenomic features associated with certain spatial structures in diverse cell types. Second, we will improve our existing predictive machine-learning models for regulatory genomics ([28-32](#_ENREF_28)) by incorporating information derived from 3D structures. For example, we will improve our existing models for gene expression prediction by taking into consideration the epigenetic (such as H3K27ac and H3K4me1) and genomic (such as p300 binding) data at their distal enhancer sites that we can infer from our 3D models. These models will reveal the significance and contribution of distal regulation in diverse cellular contexts and cell types.

*3.2. System-level functional interpretation of the 3D genome.* To capture the system-level impact of functional annotations on the 3D genome, we will also make use of interaction networks. These network models enable us to study functionally annotated DNA-DNA interactions at the systems level to detect critical components of a system and to identify subcomponents that are prone to system-level failures. To model ChIA-PET datasets in the form of networks, we have studied the DNA-DNA network architectures mediated by Pol2 ([33](#_ENREF_33)). We have further developed a web-accessible software tool, ChIA-PET Network Inference and Query Tool (CNIQT), to build DNA-DNA interaction networks and annotate these networks with functional genomics data, including epigenetic datasets, NHGRI GWAS catalog and Gene Ontology terms. We confirmed that CNIQT identifies the most critical network nodes and edges based on the network topology measures, such as highly connected enhancer nodes that harbor disease-causing variants ([34](#_ENREF_34)). CNIQT enables us to study the system-level impact of critical functional domains (e.g., super-stretch enhancers ([34](#_ENREF_34)) and broad domains ([35](#_ENREF_35))) and to identify the cascade of regulatory events that might be caused by disruptions of these sites. We will expand this interaction network analysis tool in two major directions: 1) inter-connecting 3D structures with functionally annotated interaction networks using the 3D coordinates inferred from our 3D models, and 2) building and visualizing interaction networks at multiple scales (macro, meso and micro) that reflect the hierarchical structure of the nucleome to study functional data at the appropriate spatial scale. For example, nuclear lamina-DNA associations take place at macro-scale ([36](#_ENREF_36)), whereas promoter-enhancer interactions take place at micro-scale ([33](#_ENREF_33)). We will construct networks at multiple scales by using coarse-grained heatmaps that represent the contact frequencies at each scale. We will also incorporate into 3D-NOME a component to simultaneously query and visualize corresponding portions of interaction networks and 3D structures at multiple scales to ease the visualization and study of functional sites both in 3D and in network models.

**OUTREACH AND EDUCATION**

The computational approaches developed here comprise an important set of tools for elucidating and comparing the 3D structures of genomes. In keeping with the overarching mission of NOFIC, we will make all code open-source and all methods available online through coordination with the 4DN organizational hub. We will also integrate the methodologies into an annual workshop hosted at JAX dedicated to 3D genomic technologies (see Dr. Litwin letter), thereby enabling us to proactively engage members of the biomedical community. As a software engineering practice, by specifying well-defined interfaces between components we will enable future extension or even replacement of any components. Minimizing the cost of modifying the code base is important to encouraging collaboration and wider community participation in developing future capabilities for 3D-NOME. These efforts will ensure that the tools we develop are disseminated for broad use.

**PROJECT TIMELINE**



**BIBLIOGRAPHY AND REFERENCES CITED**

1. Dada JO, Mendes P. Multi-scale modelling and simulation in systems biology. Integr Biol (Camb). 2011;3(2):86-96. Epub 2011/01/08. doi: 10.1039/c0ib00075b. PubMed PMID: 21212881.

2. Martins ML, Ferreira SC, Jr., Viela MJ. Multiscale models for biological systems. Current Opinion in Colloid & Interface Science. 2010;15(1-2):18-23. doi: 10.1016/j.cocis.2009.04.004.

3. Knijnenburg TA, Ramsey SA, Berman BP, Kennedy KA, Smit AF, Wessels LF, et al. Multiscale representation of genomic signals. Nat Methods. 2014;11(6):689-94. doi: 10.1038/nmeth.2924. PubMed PMID: 24727652; PubMed Central PMCID: PMC4040162.

4. Sharma S, Ding F, Dokholyan NV. Multiscale modeling of nucleosome dynamics. Biophys J. 2007;92(5):1457-70. doi: 10.1529/biophysj.106.094805. PubMed PMID: 17142268; PubMed Central PMCID: PMC1796817.

5. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289-93. doi: 10.1126/science.1181369. PubMed PMID: 19815776; PubMed Central PMCID: PMC2858594.

6. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013;502(7469):59-64. doi: 10.1038/nature12593. PubMed PMID: 24067610; PubMed Central PMCID: PMC3869051.

7. Zhang Z, Li G, Toh KC, Sung WK. 3D chromosome modeling with semi-definite programming and Hi-C data. J Comput Biol. 2013;20(11):831-46. doi: 10.1089/cmb.2013.0076. PubMed PMID: 24195706.

8. Kruskal JB. Nonmetric multidimensional scaling: A numerical method. Psychometrika. 1964;29(2):115-29. doi: 10.1007/BF02289694.

9. Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J. 3D genome reconstruction from chromosomal contacts. Nat Methods. 2014;11(11):1141-3. doi: 10.1038/nmeth.3104. PubMed PMID: 25240436.

10. Zhang Y. Progress and challenges in protein structure prediction. Current opinion in structural biology. 2008;18(3):342-8. doi: 10.1016/j.sbi.2008.02.004. PubMed PMID: PMC2680823.

11. Capra JA, Singh M. Characterization and prediction of residues determining protein functional specificity. Bioinformatics. 2008;24(13):1473-80. doi: 10.1093/bioinformatics/btn214. PubMed Central PMCID: PMC2718669.

12. Varoquaux N, Ay F, Noble WS, Vert JP. A statistical approach for inferring the 3D structure of the genome. Bioinformatics. 2014;30(12):i26-33. doi: 10.1093/bioinformatics/btu268. PubMed PMID: 24931992; PubMed Central PMCID: PMC4229903.

13. Zimmermann H-J. Fuzzy Set Theory and Its Applications. 2 ed: Springer Science+Business Media, LLC; 1991.

14. Snook I. The Langevin and generalised Langevin approach to the dynamics of atomic, polymeric and colloidal systems. 1st ed. Amsterdam ; Boston: Elsevier; 2007. xvi, 303 p. p.

15. Potrzebowski WB, J. M. MinkoFit3D: modeling of macromolecular structures into density maps by 3D Minkowski sum analysis. Proceedings of the 7th Structural Bioinformatics and Computational Biophysics Meeting. 2011.

16. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallographica Section A. 1976;32(5):922-3. doi: doi:10.1107/S0567739476001873.

17. Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. Proceedings of the National Academy of Sciences. 2012;109(40):16173-8. doi: 10.1073/pnas.1204799109. PubMed PMID: 22988072; PubMed Central PMCID: PMC3479593.

18. Doyle B, Fudenberg G, Imakaev M, Mirny LA. Chromatin Loops as Allosteric Modulators of Enhancer-Promoter Interactions. PLoS Comput Biol. 2014;10(10):e1003867. doi: 10.1371/journal.pcbi.1003867. PubMed PMID: 25340767; PubMed Central PMCID: PMC4207457.

19. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. J Chem Theory Comput. 2013;9(1):461-9. doi: 10.1021/ct300857j. PubMed PMID: 23316124; PubMed Central PMCID: PMC3539733.

20. Gennes PGd. Scaling concepts in polymer physics. Ithaca, N.Y.: Cornell University Press; 1979. 324 p. p.

21. Chuang J, Grosberg AY, Tanaka T. Topological repulsion between polymer globules. The Journal of Chemical Physics. 2000;112(14):6434-42. doi: 10.1063/1.481275.

22. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics. 2004;20(17):3289-91. doi: 10.1093/bioinformatics/bth378. PubMed PMID: 15217809.

23. Faeder J, Blinov M, Hlavacek W. Rule-Based Modeling of Biochemical Systems with BioNetGen. In: Maly IV, editor. Systems Biology: Humana Press; 2009. p. 113-67.

24. Blinov ML, Ruebenacker O, Moraru II. Complexity and modularity of intracellular networks: a systematic approach for modelling and simulation. Systems Biology, IET. 2008;2(5):363-8. doi: 10.1049/iet-syb:20080092. PubMed PMID: 19045831; PubMed Central PMCID: PMC2675916.

25. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636-40. doi: 10.1126/science.1105136. PubMed PMID: 15499007.

26. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8. doi: 10.1038/nbt1010-1045. PubMed PMID: 20944595; PubMed Central PMCID: PMC3607281.

27. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012;30(3):224-6. doi: 10.1038/nbt.2153. PubMed PMID: 22398613.

28. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res. 2012;22(9):1658-67. doi: 10.1101/gr.136838.111. PubMed PMID: 22955978; PubMed Central PMCID: PMC3431483.

29. Cheng C, Gerstein M. Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells. Nucleic Acids Res. 2012;40(2):553-68. doi: 10.1093/nar/gkr752. PubMed PMID: 21926158; PubMed Central PMCID: PMC3258143.

30. Cheng C, Min R, Gerstein M. TIP: a probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles. Bioinformatics. 2011;27(23):3221-7. doi: 10.1093/bioinformatics/btr552. PubMed PMID: 22039215; PubMed Central PMCID: PMC3223362.

31. Cheng C, Yan KK, Hwang W, Qian J, Bhardwaj N, Rozowsky J, et al. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. PLoS Comput Biol. 2011;7(11):e1002190. doi: 10.1371/journal.pcbi.1002190. PubMed PMID: 22125477; PubMed Central PMCID: PMC3219617.

32. Yu H, Zhu X, Greenbaum D, Karro J, Gerstein M. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. Nucleic Acids Res. 2004;32(1):328-37. doi: 10.1093/nar/gkh164. PubMed PMID: 14724320; PubMed Central PMCID: PMC373274.

33. Sandhu KS, Li G, Poh HM, Quek YL, Sia YY, Peh SQ, et al. Large-scale functional organization of long-range chromatin interaction networks. Cell Rep. 2012;2(5):1207-19. doi: 10.1016/j.celrep.2012.09.022. PubMed PMID: 23103170; PubMed Central PMCID: PMC4181841.

34. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47. doi: 10.1016/j.cell.2013.09.053. PubMed PMID: 24119843; PubMed Central PMCID: PMC3841062.

35. Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell. 2014;158(3):673-88. doi: 10.1016/j.cell.2014.06.027. PubMed PMID: 25083876; PubMed Central PMCID: PMC4137894.

36. Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, et al. Single-cell dynamics of genome-nuclear lamina interactions. Cell. 2013;153(1):178-92. doi: 10.1016/j.cell.2013.02.028. PubMed PMID: 23523135.