**OVERALL COMPONENT**

**Title: Nucleome Positioning System for Spatiotemporal Genome Organization and Regulation**

**PROJECT SUMMARY**

This proposal seeks to fulfill a community need for a comprehensive, high-resolution genome-mapping platform that will enable investigation of the structural, functional and spatiotemporal organization of the human genome. Our ultimate goal is to deliver complex chromatin interaction network maps in the context of 3D genome structures from which the dynamics of individual genomic elements can be monitored and referenced. Here, we propose to develop a **Nucleome Positioning System (NPS)**—comprised of 1) a robust genome-wide mapping technology platform, 2) advanced computational modeling algorithms and 3) state-of-the-art nuclear imaging methods—that will allow users community-wide to uncover the regulatory functions of 3D genome organization in human cells. NPS will be based upon the ChIA-PET genomic sequencing method, enhanced by process optimizations—i.e., microfluidic-based miniaturization and Tn5-transposase-based library preparation—to facilitate the study of chromatin interactions mediated by protein factors across a broader range of human cell types **(Aim 1).** We will also optimize RICh-PET for the comprehensive mapping of chromatin interactions mediated by non-coding RNAs **(Aim 1)**. The high-quality mapping data generated through these optimization efforts will be analyzed by a new computational platform (Three-Dimensional Nucleome Modeling Engine, or 3D-NOME) that makes use of hierarchical multi-scaling to model 3D genome structures **(Aim 2)**. We will also complement the 3D modeling with transcriptome, epigenome and SNP data associated with genetic diseases (GWAS) to provide functional annotation to the various structural units **(Aim 2)**. We will continue by developing strategies to validate the nucleome geometry predicted by 3D-NOME both structurally, using new nuclear imaging technologies, and functionally, using cutting-edge genome- and epigenome-editing approaches, in both human cell lines and mouse models **(Aim 3)**. Finally, we will implement NPS to generate pilot 3D genome maps from a wide range of human cell lines and primary immune cells sorted from whole blood, to elucidate the spatiotemporal dynamics of human genome organization over major developmental and hematopoietic cell lineages, as well as, among differentiating lymphocytes involved in the immune response **(Aim 4)**. Together, these efforts will yield a powerful set of sophisticated, high-quality tools and mapping data for the larger research community, and will help establish the standards for future 3D/4D nucleome studies. They will also provide insights into the broad mechanisms that organize the structure and regulate the function of the human genome, as well as the specific mechanisms by which immune responses are regulated at the nuclear level.

**PUBLIC HEALTH RELEVANCE**

It is becoming increasingly recognized that higher-order genome organization crucially influences gene regulation, cell function and ultimately human health. However, the ability to investigate these relationships will depend on technological advances that enable a more precise, integrated view of genome structure and function. This proposal seeks to address this challenge through the development of a Nucleome Positioning System that will provide a powerful platform for studying genomic structure in space (3D) and time (4D).

**SPECIFIC AIMS**

This proposal addresses outstanding technological and conceptual challenges that hinder the broad scientific community in the study of higher-order nuclear architecture and its impact on human health and disease. Current technologies for mapping nuclear structure, including ChIA-PET([1](#_ENREF_1), [2](#_ENREF_2)) and HiC([3](#_ENREF_3)), generate reliable 2D interaction data between promoters, enhancers, and other regulatory elements that are the starting point for inferring 3D genome organization computationally. However, the accuracy of these predictions is unclear. Microscopy-based techniques, e.g., FISH([4](#_ENREF_4)), can delineate broad structural features of nuclei and chromosomes, but lack the resolution of genomic sequencing methods to reveal detailed structural information or function. More recent developments (e.g., CRISPR) enable labeling of specific chromosomal regions in live cells, providing new opportunities for integrated structural and functional nucleome studies([5](#_ENREF_5)).

Our ultimate goal in this proposal is to deliver a Nucleome Positioning System (NPS) for the generation of complex chromatin interaction network maps in the context of 3D genome structures from which the dynamics of individual genomic elements can be monitored and referenced. NPS will comprise 1) a high-throughput mapping platform based on the established ChIA-PET method, enhanced through process optimizations that will expand its versatility to a broader range of cell types **(Aim 1)**, combined with 2) a sophisticated computational modeling engine, 3D-NOME, that will use hierarchical multi-scaling approaches to analyze the nuclear structure data and build spatial 3D models **(Aim 2)**. Our publications and preliminary results herein show that ChIA-PET uncovers functional interactions mediated by a protein of interest, non-enriched chromatin contacts reflecting large-scale chromosome topology, and haplotype-specific interactions, making ChIA-PET an ideal genomic sequencing technology with which to generate chromatin contact maps and to define overall nuclear architecture and its role in gene expression. ChIA-PET maps will be validated structurally using nuclear imaging technologies and in functional studies *via* genome editing in human cell lines and mouse models, and 3D-NOME will be trained and improved by nuclear imaging data **(Aim 3)**. The NPS platform will then be applied to generate high-quality pilot 3D maps in major human cell lineages and in primary human immune subtypes during differentiation and/or in response to various stimuli **(Aim 4)**.

To achieve these goals, we assembled an inter-disciplinary team of accomplished investigators and experts in genome technology, computational biology and immunology, who will work closely with each other and the broader 4D Nucleome network to help develop community standards and metrics for data generation, storage and analysis, and visualization. We expect our efforts will yield an integrated, powerful, and potentially transformative approach poised to uncover the fundamental guidelines governing genome structure and function in space (3D) and time (4D). In turn, this effort will enable the broader scientific community to address currently inaccessible questions related to higher-order nuclear function in health and disease. The specific aims of the proposal are to:

**Aim 1. Develop and improve technologies for high-resolution and haplotype-specific 3D mapping**. As described in the *Mapping Technology Development* component, we will use advanced molecular biology and microfluidic approaches([6](#_ENREF_6), [7](#_ENREF_7)) to significantly improve the efficiency of ChIA-PET and RICh-PET methods. Specifically, we will incorporate Tn5-based transposition([8](#_ENREF_8)) to increase the efficiency of library preparation, and apply microfluidics to miniaturize and automate the process. Our ultimate goal is to reduce the input sample cell requirements from the current 108 down to 100K, which will allow for a variety of biological questions that rely on the analysis of developmental compartments using limited numbers of cells.

**Aim 2. Refine the 3D-NOME system to visualize, model, and annotate the human genome.** 3D-NOME is an integrative and comprehensive data-analysis platform we recently developed to study nuclear architecture. In the *Data Analysis and Modeling* component, we will further develop the 3D-NOME to model the nuclear structures of the human genome in different cell types at high resolution. We will also integrate transcriptome, epigenome, and genetic datasets (GWAS) to further refine these analyses.

**Aim 3. Structurally and functionally validate 3D mapping and computational technologies.** For the *Biological Validation* component, we will structurally validate the predicted topological features derived from ChIA-PET using nuclear imaging and genome editing. We will use improved 3D-FISH([4](#_ENREF_4), [9-12](#_ENREF_9)) and develop live cell imaging tools to verify and refine predicted 3D structures and monitor dynamic positions of key genomic units in the nucleome. For functional validation, we will use TALEN and CRISPR([5](#_ENREF_5), [13](#_ENREF_13), [14](#_ENREF_14)) technologies to create knockin and knockout strains in human cell lines and then at the organismal level using mouse models.

**Aim 4. Elucidate the spatiotemporal regulation of 3D human genome organization and function.** To demonstrate NPS and produce valuable 3D genome data for the research community, in the *Data Generation* component, we will use the refined methodologies developed in the first three aims to produce high-quality data and build pilot 3D maps of the human genome from both cell lines and primary immune cells.

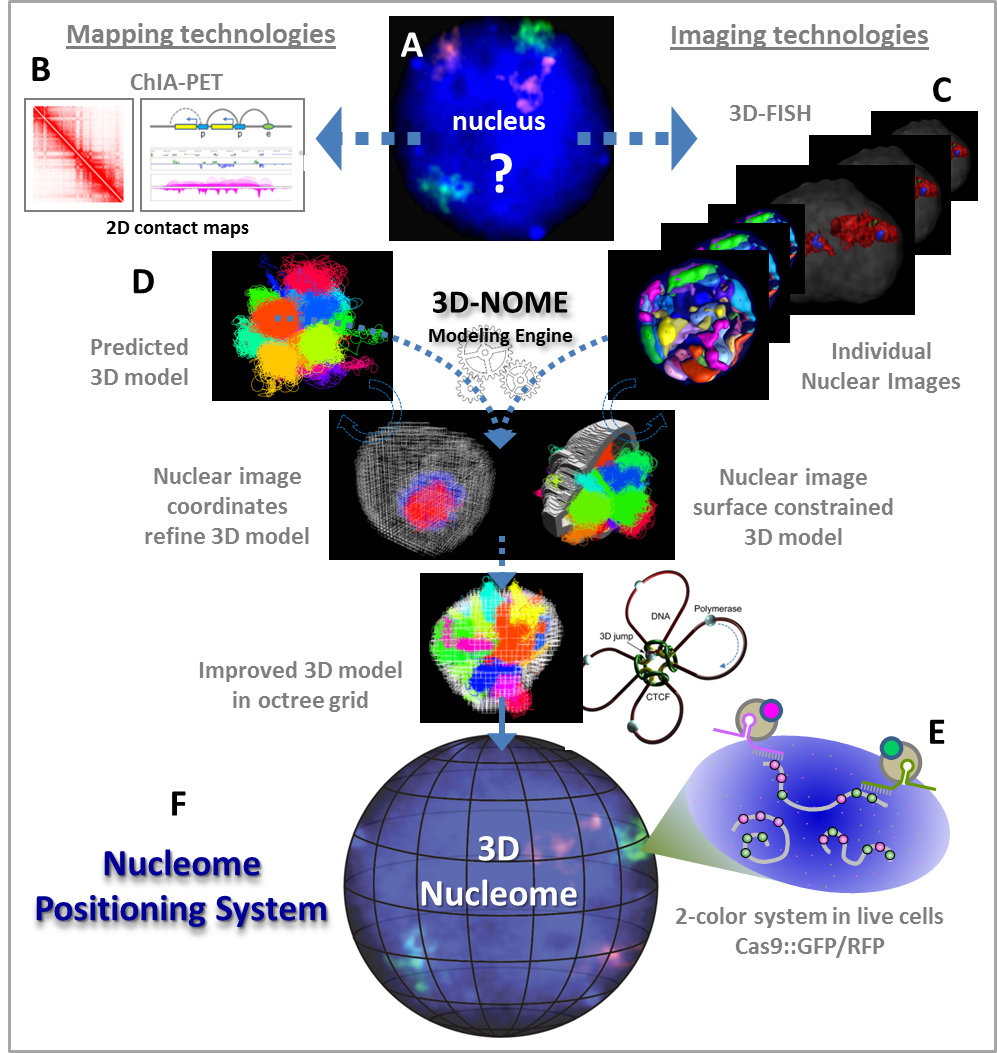
**RESEARCH STRATEGY**

**SIGNIFICANCE**

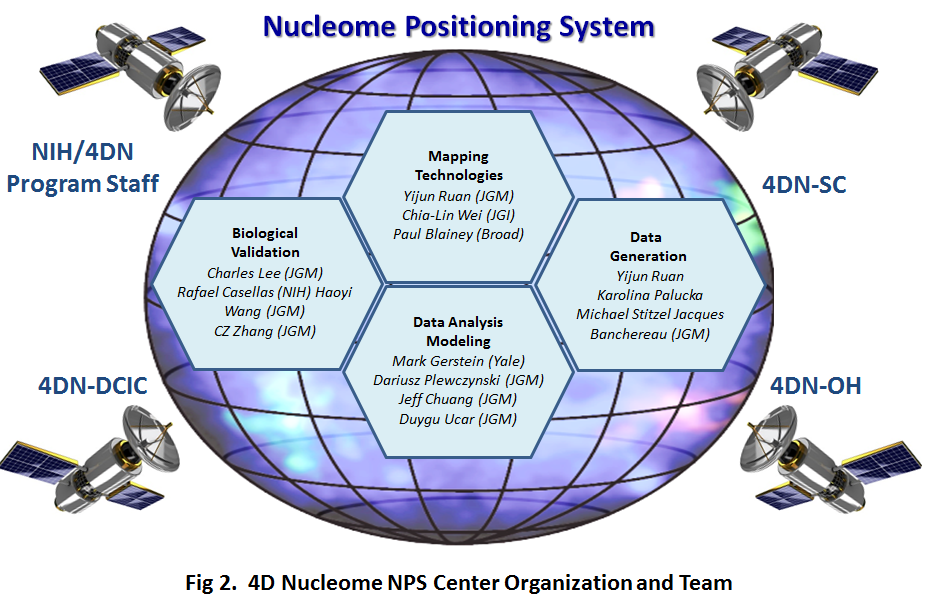
***Opportunities and challenges*.** Current advances in genomic and epigenomic analyses have identified large numbers of regulatory elements, including enhancers and non-coding RNAs (ncRNAs) ([15](#_ENREF_15" \o "Consortium, 2012 #22), [16](#_ENREF_16" \o "Djebali, 2012 #19)). Our work ([1](#_ENREF_1" \o "Fullwood, 2009 #1), [2](#_ENREF_2" \o "Li, 2012 #2)) as well as that of others ([17](#_ENREF_17" \o "Kieffer-Kwon, 2013 #3), [18](#_ENREF_18" \o "Zhang, 2013 #4)) has shown that these elements drive gene expression by interacting with active promoters over long distances. These observations underscore the importance of high-order chromatin structure in basic nuclear functions. However, the full impact of nuclear architecture in human health and disease remains to be determined. Currently available genomic technologies for mapping nuclear structure, including ChIA-PET and HiC, can generate reliable two-dimensional (2D) contact maps between promoters, enhancers, and other regulatory elements. These data are the starting point for inferring 3D genome organization computationally. However, how accurately these predictions reflect 3D nuclear organization is unclear. Ideally, one would want to complement these studies with microscopy techniques; e.g., by labeling chromosomes with fluorescent probes to provide structural information of nuclei. Early work using FISH established the concept that chromosomes occupy particular nuclear domains dubbed chromosome territories (CTs) ([4](#_ENREF_4" \o "Cremer, 2008 #8)). However, the approach lacks the resolution and specificity of genomic sequencing methods to reveal detailed structural information or function. To overcome these challenges, more recent developments (e.g., CRISPR ([5](#_ENREF_5" \o "Chen, 2013 #13))) have enabled labeling of specific chromosomal regions in live cells. Such techniques provide new opportunities for structural and functional nucleome studies.

***Overall vision.*** We seek to fill a community need for a comprehensive, high-resolution genome-mapping platform that will enable investigation of the structural, functional and spatiotemporal organization of the human genome. We will accomplish this by establishing a data production center devoted to the development of a **Nucleome Positioning System (NPS) (Fig. 1)** that will enable a richer understanding of human 3D genome organization, regulatory functions and dynamics. NPS will comprise a robust genome-wide mapping technology, state-of-the-art imaging and advanced computational modeling. High-quality mapping data will be used to computationally model global genome architecture involving functional regulatory elements. The predicted nucleome geometry will be validated structurally by imaging analysis. Topological coordinates of nuclear images will be extracted to refine 3D modeling iteratively, and the prediction program will be further improved by training using the imaging data. Suggested topological functions will be validated by genome-editing approaches, while the dynamics of individual genomic units (e.g. regulatory domains, TADs, and chromosomes) will be monitored in individual live cells. Finally, we will apply NPS to generate pilot 3D maps in both human cell lines and primary immune cells as a function of differentiation and activation, to probe the fundamental principles of genome organization and plasticity.

Figure 1. The Nucleome Positioning System (NPS). (A) The genome in the nuclear space is fluidic and dynamic, with many unknowns. (B) ChIA-PET is a genomic mapping technology that generates 2D data for inferring 3D models. (C) 3D-FISH is a nuclear imaging technology that visualizes the physical existence of chromosome inside nuclei. (D) 3D-NOME is a computational platform that predicts 3D models from ChIA-PET data, extracts topological coordinates from nuclear images, and uses it to refine 3D models of entire nucleome or specific topological structures reiteratively. (E) A two-color labeling system that works in live cells to monitor positions of specific genomic units (chromosome, topological domain, compartment, specific loci). (F) Together, the system constitutes NPS.



***Mission and organization of the center.*** The primary mission of the center is to develop the NPS platform that will, in turn, provide valuable tools, technologies, computational programs and data for the community to explore and apply to a wide variety of biological questions. Over the project period we will generate data, test and refine individual technologies and integrate these technologies into a cohesive NPS platform. We will accomplish this through four interactive center elements **(Fig. 2)**: **i)** a **Mapping component** **(Aim 1)** that seeks to merge the ChIA-PET and RICh-PET techniques with novel library generation methods and microfluidics-based process miniaturization to create a sequencing platform with unprecedented capabilities; **ii)** a **Data Analysis and Modeling component** **(Aim 2)** that will incorporate cutting-edge mathematical principles and computational capabilities to infer 3D genome organization, its dynamics and functional relevance; **iii)** a **Validation Component** **(Aim 3)** that will develop new nuclear imaging and genome-editing approaches to validate the mapping data and computational models generated in ii); and **iv)** a **Data Generation component (Aim 4)** to create pilot 3D maps of genome structure.

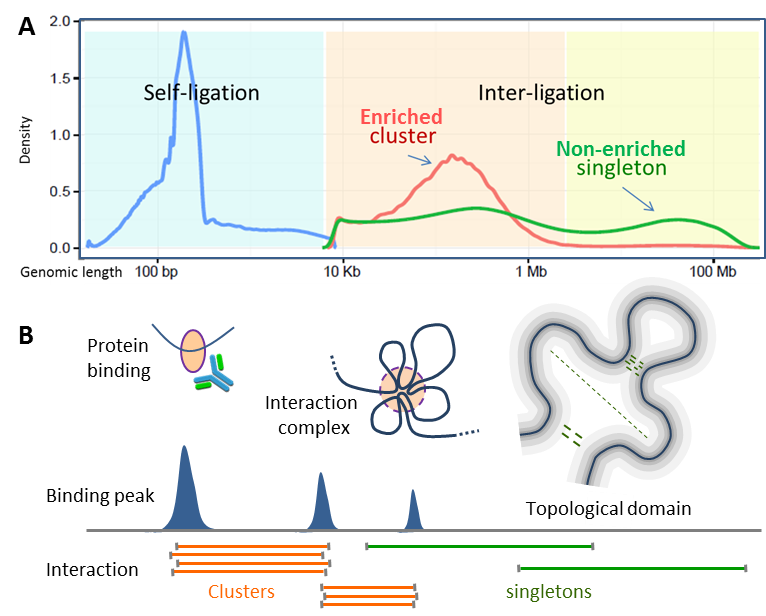
**THE TEAM.** We have assembled an interdisciplinary team (**Fig. 2)** of highly accomplished scientists whose complementary expertise in genome technologies, biophysics, computational biology and immunology will synergize to bring about the program’s goals. ***Dr. Yijun Ruan***, PI, is a pioneer in the development of high-throughput technologies to identify long-range chromatin interactions and explore 3D genome biology, including the ChIA-PET and RICh-PET methods that form the basis for this proposal. Other key individuals bring significant expertise in epigenomics *(Stitzel, Ucar, Wei);* cytogenetics *(Lee, Zhang);* genome editing *(Wang, Casellas);* biophysics and computational biology **(***Gerstein, Plewczynski, Chuang);* immunology *(Palucka, Banchereau, Casellas);*and nano-engineering and single-cell genomics *(Blainey).* Our team members have a history of productive collaborations and rich experience with large consortium efforts (*Wei, Gerstein, Ruan***,** ENCODE*; Lee, Plewczynski, Gerstein,* 1000 Genomes Project). Team members within and outside The Jackson Laboratory have a strong record of sharing protocols, tools and data broadly with the academic research community without license or research restriction. In keeping with the goals of NOFIC, the team will work closely with each other and the broader 4D Nucleome (4DN) network to help develop community standards and metrics for data generation, storage and analysis, and visualization. This interactive, diverse team is thus ideally positioned to accomplish the proposed work.

**INNOVATION**

The originality of our center lies in the development and integration of cutting-edge technologies—i.e., microfluidics, genome mapping, nuclear imaging, computational modeling—to provide a global view of human nuclear architecture and function. Here, we summarize the key innovations of our approach. **1)** The ChIA-PET method is unique among 3C technologies in the multiplicity of rich data it can yield from a single experiment, i.e., i) specific factor binding profiles, ii) long-range interactions involving the immunoprecipitated factor, iii) chromatin interactions that do not involve the immunoprecipitated factor, and iv) haplotype specificity. **2)** RICh-PET is a novel method that identifies chromatin interactions in an unbiased all-to-all manner. The application of RICh-PET will reveal the role of RNA in nuclear architecture. **3)** Our application of microfluidics will popularize the use of ChIA-PET and RICh-PET, which so far have been used by very few laboratories. **4)** We will further develop and refine 3D-NOME, a comprehensive computational engine that uses a hierarchical, multi-scale approach to spatially model, visualize, and annotate the human nucleome. **5)** We are developing nuclear imaging methods, including 3D imaging of open chromatin regions and a multi-color imaging system for live cells, that will provide essential structural insight to validate our computational models and train 3D-NOME. **6)** We will functionally validate the predicted chromosome interactions via epigenome editing not only in cell lines, but also at the organismal level by creating transgenic mice with multi-color TALE/dCAs9::fluorophores. This will further enable studies of the dynamics of chromosome behavior in primary cells. **7)** We will merge these datasets to generate pilot 3D maps in human cell lines and, for the first time, primary human cells. **The result will be an unparalleled system and set of tools for high-dimensional, high-resolution analysis of genome organization.**

**APPROACH**

***Overall strategy****.* Our ultimate goal is to develop a NPS in which the dynamics of regulatory elements can be monitored in the complex 3D maps of chromatin interaction network during cell differentiation. To achieve this goal, we are establishing a set of microfluidic and high-throughput technologies to generate multi-scale, high-resolution and haplotype-specific chromatin interaction data that we can use to model the structure of the human genome, even for tissues where cell numbers are limited. Nuclear architecture will be mapped by ChIA-PET and RICh-PET. We summarize our progress followed by the overarching goals of each Aim below, and refer the reader to the associated research plan for details.



**Figure 3. Characteristics of ChIA-PET data.** **(A)** ChIA-PET data include self-ligation and inter-ligation products. Self-ligation data are mapped in short range (100-500bp). Enriched interaction cluster data mostly in the range of 10Kb-2Mb and peaked at 200Kb. Non-enriched singleton data are scattered flat in all range. **(B)** Self-ligation data are used to identify protein binding peaks, enriched chromatin clusters are to detect specific chromatin interactions mediated by protein factor of interest, and singleton data are to plot large-scale topological domains which suggest relative neighborhood positions of chromatin segments.

***ChIA-PET: powerful 3D genomic technologies that form the basis for NPS.*** ChIA-PET is a unique among the 3C techniques in that it involves an immunoprecipitation step that enriches for functional interactions and provides a means to detect reproducible interactions between promoters, enhancers and insulators **(Fig. 3)**. Our recent analyses showed that, in addition to protein-mediated interactions, ChIA-PET also uncovers broad chromosome conformation characteristics similar to HiC **(**see **Fig. 5, below).** Furthermore, we recently extended the read-length of ChIA-PET tags from 20b to 150bp. This improvement significantly increases tag-sequence coverage, thereby permitting detection of allele-specific interactions if haplotype information is available. In GM12878 and HeLa, we identified haplotype-specific chromatin interactions by CTCF and RNAPol2. We further observed that 95% of SNP-marked *cis*-interactions (i.e., intra-chromosomal) data, including singletons, are associated with the same haplotype (data not shown), suggesting that the majority of ChIA-PET-identified interactions take place in *cis*. This observation is in agreement with HiC data ([19](#_ENREF_19" \o "Rao, 2014 #16), [20](#_ENREF_20" \o "Selvaraj, 2013 #18)) and consistent with the CT concept ([3](#_ENREF_3" \o "Lieberman-Aiden, 2009 #17)), whereby each chromosome tends to fold onto itself. Based on this property, we can study haplotype phasing and homologous-specific chromosome folding of all 46 chromosomes.

With the current ChIA-PET protocol, we routinely generate high-complexity libraries using one lane of a Hiseq run to generate 50-100 million high-quality mapping reads per library, which provides sufficient depth of sequencing data for detecting binding peaks, specific chromatin interactions between binding sites (i.e., mapping clusters with multiple PET counts) and non-enriched singleton data. A triplicate RNAPol2 ChIA-PET dataset confirms high reproducibility at both global and specific segment levels **(Fig. 4).**

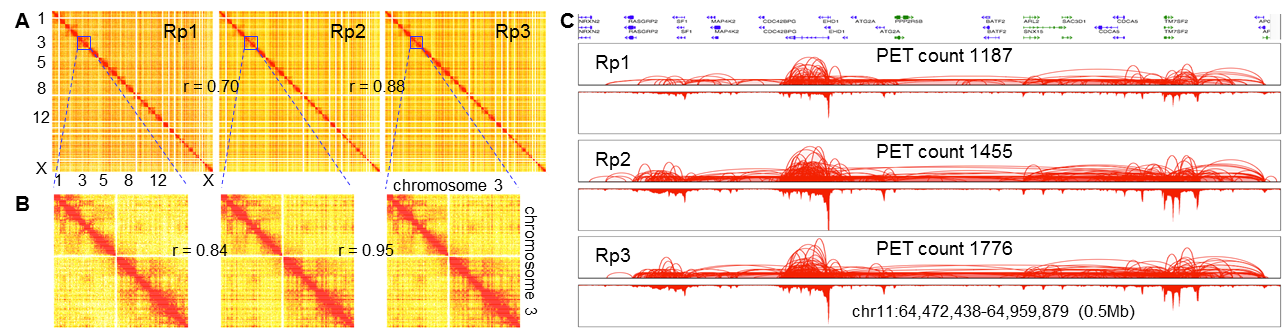


Figure 4. Triplicate RNAPol2 ChIA-PET data from GM12878 cells. (A) Heat maps of genome-wide interaction contact frequency (Rp1 quality reads: 51,505,047; Rp2: 96,132,545; Rp3: 85,500,327). (B) Heat maps of chr3 interaction contact frequency. Correlation r-value calculated by Pearson’s correlation coefficient. (C) Screenshot showing mapping of looping (top) and binding (bottom) by RNAPol2 in a 0.5Mb region of chr11.

***Comparative analysis of ChIA-PET and HiC confirms high ChIA-PET performance.*** In a recent study, Rao and colleagues used a new version of HiC, *in situ* HiC, to generate 4.9 billion contact reads from GM12878 cells, achieving kilobase resolution and finding that the majority of interaction loops were associated with CTCF binding sites ([19](#_ENREF_19" \o "Rao, 2014 #16)). As we too generated CTCF ChIA-PET data from GM12878 cells, we had a perfect opportunity to compare HiC and ChIA-PET performance. Using the same data analysis method, we generated contact heatmaps of our CTCF ChIA-PET data in chr7 and compared these to the *in situ* HiC chr7 data **(Fig. 5).** The two sets of heatmaps displayed very similar patterns, suggesting a high degree of fidelity among the captured chromatin interactions. In addition, within the same library data, ChIA-PET analysis identified specific CTCF loops with binding site resolution (50-100 bp). Furthermore, it is important to note that 4.9 billion mapping reads were necessary for these *in situ* HiC data to reach kilobase resolution, whereas only one Hiseq (56M PET reads) or Miseq run (11M PET reads) was needed to achieve the same outcomes using ChIA-PET, and with 10-fold higher resolution and 73-fold (Hiseq) or 445-fold (Miseq) less sequencing reads. These findings suggest that ChIA-PET offers an unparalleled combination of specificity, resolution and performance.

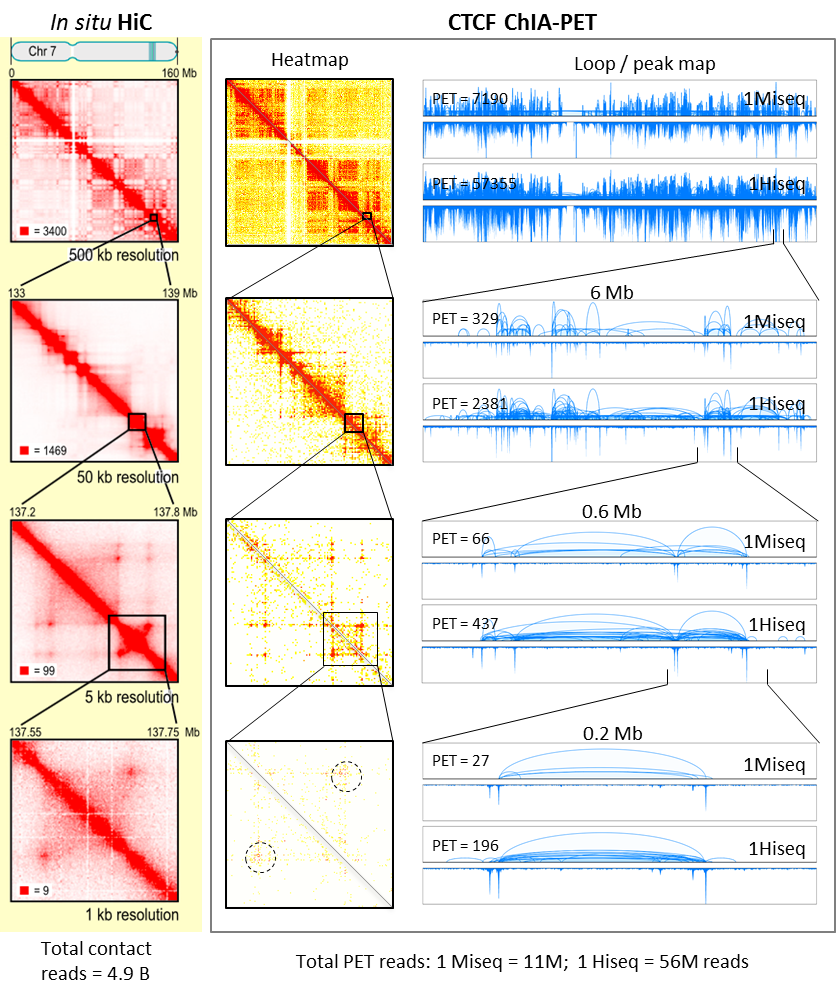
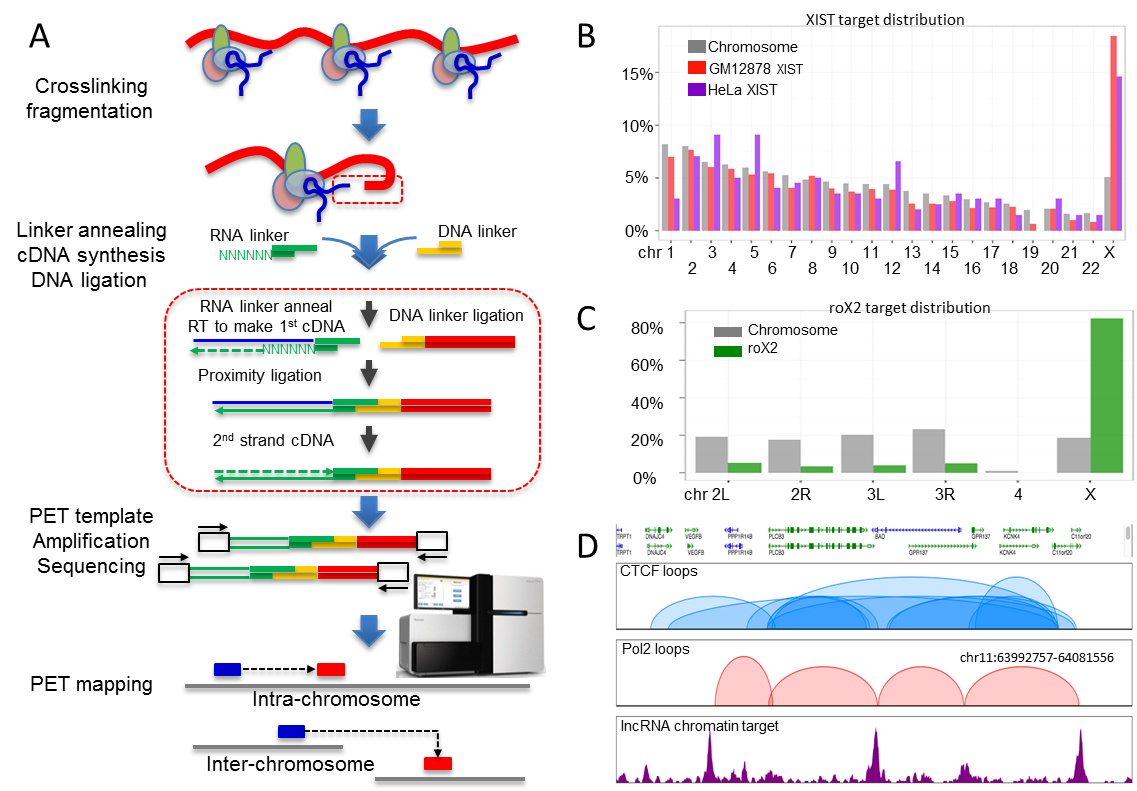


Figure 5. Comparison between HiC and ChIA-PET datasets. (Left) HiC contact heatmap of Chr7 (Rao et al 2014, Figure 1). (Middle) CTCF ChIA-PET contact heatmap of Chr7 and matched regions with the HiC data. (Right) CTCF ChIA-PET loop/peak map view. For each dataset (1Miseq, 1Hiseq), loop view is on top, peak view is bottom; height indicates the intensity of looping and binding. 1Miseq, library from one Miseq run; 1Hiseq, data from one Hiseq run. PET counts: numbers of interaction PETs detected in the given region.

***RICh-PET, a novel technology to map RNA-chromatin interactions.*** Recent large-scale sequencing efforts have identified a number of long, non-coding RNAs (lncRNAs) ([16](#_ENREF_16" \o "Djebali, 2012 #19)). Their ubiquity, evolutionary conservation and nuclear localization all imply important roles in genomic regulation, but their precise functions are a puzzle. Recent studies found that lncRNAs are involved in regulating ER-mediated transcription programs ([21](#_ENREF_21" \o "Li, 2013 #7)). However, current methods to investigate these questions are mostly limited to a single RNA or target at a time ([22-24](#_ENREF_22" \o "Chu, 2011 #25)), and most of the inferred lncRNA functions, with the exception of a few well-studied ones, are based on an indirect “guilt by association” ([25-27](#_ENREF_25)). This knowledge gap is largely due to the lack of efficient technologies to study lncRNA functions. We suggest that lncRNAs, having an epigenetic regulatory role, must interact with chromatin at specific chromosomal locations directly or indirectly through protein intermediates.

To overcome current limitations, we developed RICh-PET, a new technology for unbiased genome-wide mapping that uses specially designed DNA oligonucleotides to convert interactive RNA molecules (such as lncRNA) into cDNAs that can then be ligated to chromatin DNA fragments that are tethered together. The RNA (cDNA) and DNA ligation products are then subjected to paired-end-tag library construction, directional sequencing and mapping to the reference genome (**Fig. 6A**), to identify lncRNAs (RNA-tag sequences) and their interacting chromatin sites (DNA-tag sequences). RICh-PET experiments using human HeLa cells, GM12878 and *Drosophila* S2 cells demonstrate that this method is highly specific (**Fig. 6B,C**) and reproducible in re-identifying known lncRNAs and uncovering large numbers of new lncRNAs and their chromatin target sites. Using RICh-PET, we have comprehensively analyzed lncRNAs and their chromatin interaction loci in two human cell lines (B-lymphocyte GM12878 and HeLa). Remarkably, many lncRNAs showed long-distance *cis-* and *trans-*acting interactions specifically to gene-promoter regions, suggesting potential regulatory functions. Furthermore many of chromatin target sites by lncRNAs are associated with chromatin loop anchor regions as detected by ChIA-PET data, indicating roles of lncRNAs in chromatin topological structures **(Fig 6D)**. Knockdown experiments by siRNA indicate that lncRNA molecules indeed modulate transcription levels, both up- and down-regulating their target genes. Further computational analysis indicated there is significant enrichment of lncRNA target genes involved in cancer processes such as cell growth, apoptosis and invasion (manuscript in reviewing). Further refinement and application of RICh-PET to study lncRNAs in cancer cells is currently supported by NCI (**1R01CA186714-01**).

**Figure 6. The RICh-PET method. (A)** Schematic of the method. **(B and C)** Bar charts showing the DNA tags connected to X-chr specific lncRNAs XIST in human and rox2 in *Drosophila* are significantly enriched in X-chromosome, thus providing the validation of specificity to the RICh-PET method. **(D)** The chromatin targeting by lncRNAs (purple peaks) are enriched in interaction anchor regions that are defined by CTCF/Pol2 ChIA-PET data.



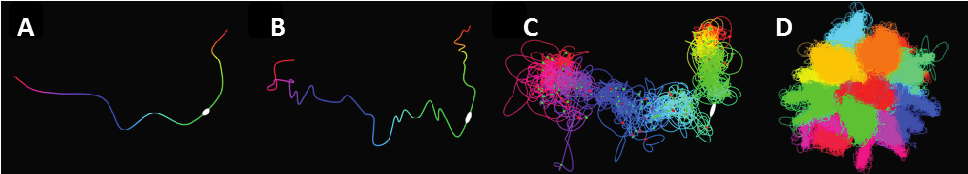
**Aim 1. Develop and improve technologies for high-resolution and haplotype-specific 3D mapping.**

***Overall strategy****.* Our ultimate goal is to develop a NPS composed of complex chromatin interaction network maps in the context of 3D genome structures, in which each genomic element can be monitored for their dynamics. To achieve this goal, we plan to establish a set of high-throughput 3D-mapping technologies to generate multi-scale, high-resolution and haplotype-specific chromatin interaction data to model the ensemble structures of human genomes. As we demonstrate above, ChIA-PET has this potential (**Figs. 3-5**). It has also been speculated that many ncRNAs, particularly lncRNAs, have regulatory functions and are presumably involved in chromatin interactions. RICh-PET was designed to identify lncRNA-chromatin interactions in an unbiased, genome-wide manner **(Fig 6)**. In this **Mapping Technology Development** component, we seek to further refine ChIA-PET and RICh-PET for generating genome-wide high-quality and haplotype-specific maps of chromatin interactions with significantly increased efficiency and robustness. We propose to accomplish this *via* two sets of improvements: i) Tn5 transposase-assisted transposition combined with a new annealing and extension step, to increase efficiency of ChIA-PET and RICh-PET library preparation, and ii) a microfluidics-based platform to miniaturize and automate the process. This will allow us to vastly reduce input cell quantity from the required 108 down to 105 or less. We will also establish the standards of chromatin interaction analysis for generating high-quality, high-resolution and haplotype-specific 3D maps of human genomes.

**Aim 2. Refine the 3D-nucleome modeling engine (3D-NOME) for modeling, visualizing and annotating the human nucleome.**

***Overall Strategy.*** Elucidating 3D human genome organization requires computational advances in data pre-processing, 3D data modeling, data visualization and functional annotation. ChIA-PET data can capture spatial information at multiple scales ranging from fine enhancer-promoter interactions to large topologically associating domain (TAD) structures that can be ordered into a hierarchical spatial model. We propose to develop novel computational models that incorporate multiple scales of the nucleome at all stages of data analysis, from data pre-processing to final data visualization. The centerpiece of our proposed work is an integrative and comprehensive data-analysis platform, namely 3D Nucleome Modeling Engine (**3D-NOME**), to hierarchically model and visualize the human nucleome. This engine will enable us to model spatial structures in 3D space at multiple scales. The 3D-NOME has three major components: 1) data pre-processing to denoise ChIA-PET datasets; 2) structural modeling *via* Multi-Dimensional Scaling (MDS) and refinement; and 3) visualization and functional annotation to interpret the 3D nucleome at multiple scales. Using 3D-NOME, we are able to rapidly generate structural models of individual chromosomes and for the complete nucleome at multiple resolutions **(Fig. 7).** In this **Data Analysis and Modeling** component, we seek to further develop 3D-NOME by advancing MDS algorithms, notably for novel nuclear imaging data, which we will use to refine and improve structural predictions; develop methods for evaluating and comparing structural ensembles across multiple conditions (e.g. different cell types), including improved dynamic modeling of critical substructures at fine scales and computationally efficient rule-based modeling at large scales; and visualize and annotate spatial structures to identify how they regulate cellular processes in health and disease states.

**Figure 7. The 3D models at various scales inferred by 3D-NOME. (A)** Chromosome X at 16Mb resolution, **(B)** at 2Mb resolution, and **(C)** at interaction anchor-level resolution (10-100kb). The white ellipsoid in each panel depicts the centromere region. The red and green dots in (C) are the binding sites of CTCF and PNAPol2, respectively. **(D)** A 3D nucleome of 23 pairs of the chromosomes of GM12878 cells.

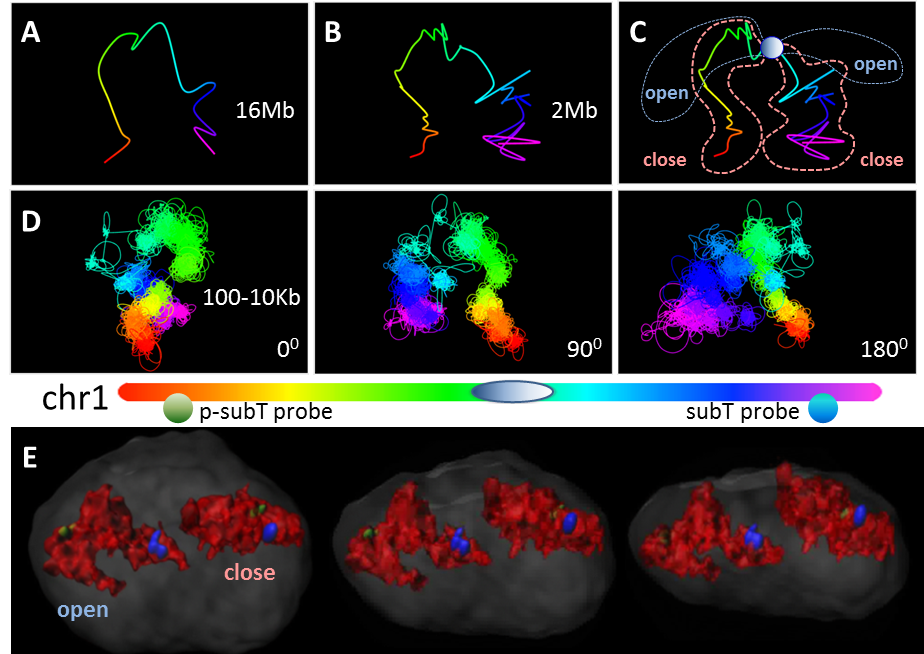


**Aim 3. Structurally and functionally validate 3D mapping and computational technologies *in vitro* and *in vivo*.**

***Overall strategy.*** This **Biological Validation** component will develop structural and functional approaches to validate our genome-wide mapping data and predicted 3D genome models. Structurally, we will apply state-of-the-art 3D-FISH to characterize the CTs of each and all chromosomes and compare 3D image features with the predicted 3D models. We will develop a new 3D-FISH method to specifically visualize open chromatin regions, thus providing the precision needed to observe active versus inactive chromosomal regions and their topological relationship in 3D. In addition to obtaining nuclear images from fixed cells, we are also developing CRISPR/Cas9 and TALE-based live-cell imaging technology to investigate chromatin/chromosome positioning in the nucleome space. Functionally, we will use TALEN and CRISPR genome-editing approaches to knock out selected chromatin-interaction anchor regions, to test the importance of these loci for 3D structure and the consequence of their deletion to transcriptional function. We are also developing CRISPR to alter epigenomic state while not physically disrupting the underlying DNA sequence, as well as to see if such epigenomic alterations affect 3D structure and transcriptional activity. In addition, we will extend our validation to the organism by creating knockout and transgenic mice of specific interest, to evaluate structural and functional impacts in all cells.

The use of nuclear imaging to validate the 3D models predicted from the genomic mapping data has yielded promising results **(Fig. 8)**. Using our GM12878 ChIA-PET data to model Chr1 at different resolutions, we identified a putative structural framework whereby the two chromosomal arms bend and extend in the same direction. Considering the mapping data were generated from millions of cells, the predicted 3D model is likely to reflect an average/representative structure. Therefore, it is possible that the ensemble Chr1 structure is plastic, fluidly adopting open and closed positions (**Fig. 8C**). To validate the predicted 3D models, we analyzed Chr1 territory by 3D-FISH. To provide orientation, we included two positioning probes (green and red) in the 3D-FISH experiment. Intriguingly, among the 50 nuclei examined so far we observed a wide range of structural variation for the two chromosome arms, from widely open to completely closed. **Fig. 8E** exemplifies the two open and closed arm conformations of chr1 territories within one nucleus. These data provide proof-of-principle of the viability of our strategy to use nuclear imaging to validate our mapping data predictions.

Figure 8. Chr1 folding dynamics in GM12878 cells. (A,B) 3D modeling of chr1 by 3D-NOME at low resolution (bin size 16Mb, 2Mb) suggests that the two arms are bent in the same direction. (C) Ensemble model of chr1 folding dynamics. (D) High resolution 3D model of chr1 with views from different rotation angles. (E) 3D-FISH images of chr1 territory. Specific position probes for the p-arm (green) and q-arm (blue) indicate that within this particular nucleus, the two arms of one copy of chr1 were bent but still “open”, whereas the two arms of the other chr1 were in “closed” form.



**Aim 4. Elucidate the spatiotemporal regulation of 3D genome organization and function in human cells.**

***Overall strategy.*** During development and refinement of our mapping technologies, we expect to produce substantial volumes of high-quality 3D genome mapping data by ChIA-PET and RICh-PET. In parallel and in concert with our technology development efforts, we will take a phased approach to producing high-quality 3D genome mapping data from established human cell lines representing major development/differentiation lineages, i.e., embryonic, endoderm, mesoderm and ectoderm cells, and primary human immune cells derived from blood. In addition, we will also generate corresponding transcriptomic and epigenomic datasets from these cell samples. Such data, along with the integration of existing ENCODE and GWAS datasets will provide thorough functional annotation to the 3D models predicted from the ChIA-PET data. In the initial phase (year 1), we will rigorously validate technical parameters such as reproducibility, sensitivity and specificity in cell lines and will produce pilot 3D genome maps. In collaboration with other 4DN groups, we will establish data quality and analysis standards. With our technology improvements in place, including the use of smaller numbers (105-106) of cells for 3D genome mapping, we will then start (year 2-3) to analyze primary hematopoietic cells of the human immune system. The mapping data generated from these experiments would reflect systematic spatiotemporal changes of 3D genomes over large, medium and fine time scales, so as to facilitate 4D nucleome studies.

**BIBLIOGRAPHY AND REFERENCES CITED**

1. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009;462(7269):58-64. doi: 10.1038/nature08497. PubMed PMID: 19890323; PubMed Central PMCID: PMC2774924.

2. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012;148(1-2):84-98. doi: 10.1016/j.cell.2011.12.014. PubMed PMID: 22265404; PubMed Central PMCID: PMC3339270.

3. 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.

4. Cremer M, Grasser F, Lanctot C, Muller S, Neusser M, Zinner R, et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. Methods in molecular biology. 2008;463:205-39. doi: 10.1007/978-1-59745-406-3\_15. PubMed PMID: 18951171.

5. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155(7):1479-91. doi: 10.1016/j.cell.2013.12.001. PubMed PMID: 24360272; PubMed Central PMCID: PMC3918502.

6. Tan SJ, Phan H, Gerry BM, Kuhn A, Hong LZ, Min Ong Y, et al. A microfluidic device for preparing next generation DNA sequencing libraries and for automating other laboratory protocols that require one or more column chromatography steps. PloS one. 2013;8(7):e64084. doi: 10.1371/journal.pone.0064084. PubMed PMID: 23894273; PubMed Central PMCID: PMC3722208.

7. Thorsen T, Maerkl SJ, Quake SR. Microfluidic large-scale integration. Science. 2002;298(5593):580-4. doi: 10.1126/science.1076996. PubMed PMID: 12351675.

8. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome biology. 2010;11(12):R119. doi: 10.1186/gb-2010-11-12-r119. PubMed PMID: 21143862; PubMed Central PMCID: PMC3046479.

9. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS biology. 2005;3(5):e157. doi: 10.1371/journal.pbio.0030157. PubMed PMID: 15839726; PubMed Central PMCID: PMC1084335.

10. Chaumeil J, Micsinai M, Skok JA. Combined immunofluorescence and DNA FISH on 3D-preserved interphase nuclei to study changes in 3D nuclear organization. Journal of visualized experiments : JoVE. 2013(72):e50087. doi: 10.3791/50087. PubMed PMID: 23407477; PubMed Central PMCID: PMC3597039.

11. Markaki Y, Smeets D, Cremer M, Schermelleh L. Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. Methods in molecular biology. 2013;950:43-64. doi: 10.1007/978-1-62703-137-0\_4. PubMed PMID: 23086869.

12. Zhao FY, Yang X, Chen DY, Ma WY, Zheng JG, Zhang XM. A method for simultaneously delineating multiple targets in 3D-FISH using limited channels, lasers, and fluorochromes. European biophysics journal : EBJ. 2014;43(1):53-8. doi: 10.1007/s00249-013-0938-5. PubMed PMID: 24305663.

13. Ma H, Reyes-Gutierrez P, Pederson T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(52):21048-53. doi: 10.1073/pnas.1319097110. PubMed PMID: 24324157; PubMed Central PMCID: PMC3876203.

14. Miyanari Y, Ziegler-Birling C, Torres-Padilla ME. Live visualization of chromatin dynamics with fluorescent TALEs. Nature structural & molecular biology. 2013;20(11):1321-4. doi: 10.1038/nsmb.2680. PubMed PMID: 24096363.

15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMC3439153.

16. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101-8. doi: 10.1038/nature11233. PubMed PMID: 22955620; PubMed Central PMCID: PMC3684276.

17. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell. 2013;155(7):1507-20. doi: 10.1016/j.cell.2013.11.039. PubMed PMID: 24360274; PubMed Central PMCID: PMC3905448.

18. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013;504(7479):306-10. doi: 10.1038/nature12716. PubMed PMID: 24213634; PubMed Central PMCID: PMC3954713.

19. Rao Suhas SP, Huntley Miriam H, Durand Neva C, Stamenova Elena K, Bochkov Ivan D, Robinson James T, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;In press. doi: 10.1016/j.cell.2014.11.021.

20. Selvaraj S, J RD, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nature biotechnology. 2013;31(12):1111-8. doi: 10.1038/nbt.2728. PubMed PMID: 24185094; PubMed Central PMCID: PMC4180835.

21. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013;498(7455):516-20. doi: 10.1038/nature12210. PubMed PMID: 23728302; PubMed Central PMCID: PMC3718886.

22. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Molecular cell. 2011;44(4):667-78. doi: 10.1016/j.molcel.2011.08.027. PubMed PMID: 21963238; PubMed Central PMCID: PMC3249421.

23. Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, Mooney SD, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome research. 2009;19(3):381-94. doi: 10.1101/gr.082503.108. PubMed PMID: 19116412; PubMed Central PMCID: PMC2661799.

24. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Molecular cell. 2010;40(6):939-53. doi: 10.1016/j.molcel.2010.12.011. PubMed PMID: 21172659; PubMed Central PMCID: PMC3021903.

25. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nature genetics. 2011;43(7):621-9. doi: 10.1038/ng.848. PubMed PMID: 21642992; PubMed Central PMCID: PMC3652667.

26. Gutschner T, Baas M, Diederichs S. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. Genome research. 2011;21(11):1944-54. doi: 10.1101/gr.122358.111. PubMed PMID: 21844124; PubMed Central PMCID: PMC3205578.

27. Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer research. 2013;73(3):1180-9. doi: 10.1158/0008-5472.CAN-12-2850. PubMed PMID: 23243023; PubMed Central PMCID: PMC3589741.