1-26-2017

The 4D Nucleome Project

Job Dekker
University of Massachusetts Medical School

Andrew S. Belmont
University of Illinois at Urbana-Champaign

Mitchell Guttman
California Institute of Technology

See next page for additional authors

Repository Citation

The 4D Nucleome Project

Authors

Job Dekker, Andrew S. Belmont, Mitchell Guttman, Victor O. Leshyk, John T. Lis, Stavros Lomvardas, Leonid A. Mirny, Clodagh C. O'Shea, Peter J. Park, Bing Ren, Joan C. Ritland Politz, Jay Shendure, and Sheng Zong

Keywords

genomics, 4D Nucleome Network, gene expression, cellular function

The 4D Nucleome Project

Job Dekker[1], Andrew S. Belmont[2], Mitchell Guttman[3], Victor O. Leshyk[4], John T. Lis[5], Stavros Lomvardas[6], Leonid A. Mirny[7], Clodagh C. O'Shea[8], Peter J. Park[9], Bing Ren[10], Joan C. Ritland Politz[11], Jay Shendure[12], Sheng Zhong[4] and the 4D Nucleome Network[13]

Correspondence: job.dekker@umassmed.edu

[1] Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Howard Hughes Medical Institute, Worcester, MA 01605
[2] Department of Cell and Developmental Biology, University of Illinois, Urbana-Champaign, IL 61801
[3] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125
[4] Department of Bioengineering, University of California San Diego, La Jolla, CA 92093
[5] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853
[6] Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10027
[7] Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139
[8] Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037
[9] Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115
[10] Ludwig Institute for Cancer Research, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, Moores Cancer Center, University of California San Diego, La Jolla CA 92093
[11] Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. Seattle, WA 98109
[12] Department of Genome Sciences, University of Washington, Seattle WA Howard Hughes Medical Institute, Seattle WA 98109
[13] A list of participating scientists is included in supplemental table 1.

Summary

The spatial organization of the genome and its dynamics contribute to gene expression and cellular function in normal development as well as in disease. Although we are increasingly well equipped to determine a genome's sequence and linear chromatin composition, studying the three-dimensional organization of the genome with high spatial and temporal resolution remains challenging. The 4D Nucleome Network aims to develop and apply approaches to map the structure and dynamics of the human and mouse genomes in space and time with the long term goal of gaining deeper mechanistic understanding of how the nucleus is organized. The project will develop and benchmark experimental and computational approaches for measuring genome conformation and nuclear organization, and investigate how these contribute to gene regulation and other genome functions. Further efforts will be directed at applying validated experimental approaches combined with biophysical modeling to generate integrated maps and quantitative models of spatial genome organization in different biological states, both in cell populations and in single cells.

Background

The human genome contains tens of thousands of genes, and a much larger number of regulatory elements, that collectively encode the program by which we develop and function as complex organisms. Large-scale studies within the last decade have extensively cataloged these components of our genome, as well as the cell types and tissues in which they are active. For instance, the ENCODE, Roadmap Epigenome, the International Human Epigenome Project and FANTOM projects [1-4] annotated tens of thousands of genes and millions of candidate regulatory elements such as enhancers, insulators and promoters in the human and mouse genomes. However, our understanding of where and how these elements contribute to gene regulation, including the mechanisms by which they exert regulatory effects on specific target genes across distances of kilobases to in some cases megabases, is decidedly incomplete.

It is increasingly appreciated that the spatial folding of chromosomes and their organization in the nucleus have profound regulatory impact on gene expression. For example, spatial proximity is necessary for enhancers to modulate transcription of their target genes (e.g. [5-7]), and clustering of chromatin near the nuclear lamina is correlated with gene silencing and the timing of DNA replication [8,9]. Meanwhile, genome-wide association studies (GWAS) have identified large numbers of disease-associated loci, and the vast majority of them are located in distal, potentially regulatory, noncoding regions (e.g. [10]). These GWAS results not only emphasize the importance of distal regulatory elements for genome functions, but also suggest an exciting opportunity to uncover fundamental mechanisms of disease through the mapping of long-range chromatin interactions and 3D genome organization. Therefore, to fully determine how the genome operates inside living cells, we need to understand not only the linear encoding of information along chromosomes, but also its 3-dimensional organization and its dynamics across time, *i.e.* the "4D nucleome". Concomitantly, we must pursue deeper knowledge of the biophysical and molecular factors that determine genome organization, and how this organization contributes to gene regulation and other nuclear activities.

Genome conformation is intimately connected to the broader organization of the nucleus. Over a century of microscopy studies have revealed that the nucleus is not a homogeneous organelle, but rather consists of distinct nuclear structures and non-chromatin bodies as well as defined chromosomal regions, such as centromeres, telomeres and insulator bodies, recognized to cluster with each other and other genomic regions to define distinct nuclear compartments [11-13]. Examples of nuclear structures include the nuclear periphery, nuclear pores and the heterochromatic compartment [14], while examples of nuclear bodies include nucleoli, nuclear speckles, paraspeckles, and Cajal and PML bodies. Chromosome conformation capture approaches [15] have yielded additional insights by characterizing chromatin folding genome-wide at kilobase-resolution [6,16-18]. These studies show that the genome is compartmentalized in active and inactive spatial compartments at the scale of the nucleus, and within each compartment folding of chromatin fibers brings together sets of

loci and regulatory elements that are otherwise separated by large genomic distances. It is possible that such interactions are at least in part related to and controlled by the formation of chromatin domains [17,19,20]. Further evidence suggests that CTCF, the cohesin complex, and many other DNA binding proteins as well as noncoding RNAs play roles in organizing chromatin domains and long-range interactions between DNA loci [17,21-26]. Altogether, these studies point to the genome being intricately organized within the nucleus, with this organization playing a critical role in gene regulation and activity.

The past decade witnessed significant technological innovation in the area of chromosome and nuclear structure analysis. Genomic approaches for mapping chromatin interactions, such as 3C, 4C, 5C, Hi-C, and ChIA-PET, are yielding genome-wide chromatin interaction maps at unprecedented resolution [18]. Live cell and super-resolution microscopic approaches, combined with application of new ways (e.g. CRISPR/Cas9 –based systems) to visualize chromosome loci and sub-nuclear structures are beginning to provide detailed views of the organization and dynamics of chromatin inside single (living) cells [27-33]. There has also been tremendous progress in modeling and analyzing chromosome structural data, producing structural models for chromosome folding which promise to provide biophysical interpretation of these often complicated datatypes [34,35]. However, despite this progress, a comprehensive understanding of the 4D nucleome is still lacking. This is partly due to the fact that different experimental cell systems and approaches are used, which together with the absence of widely shared benchmarks for assay performance have led to observations that cannot be directly compared. Additionally, we currently have limited ability to integrate different datatypes (e.g. chromatin interaction data and imaging-based distance measurements) and lack approaches that can measure and account for cell-to-cell variability in chromosome and nuclear organization. Finally, we continue to lack mechanistic insights into the relationships between chromosome conformation and nuclear processes including transcription, DNA replication, and chromosome segregation. We believe these major gaps can be addressed by a highly synergistic, multidisciplinary and integrated approach in which groups with different expertise and knowledge, ranging from imaging and genomics to computer science and physics, work closely together to study common cell systems using complementary methods.

Overview of the goals and strategy of the project

The 4D Nucleome (4DN) Network aims to develop a set of validated approaches to map the structures and dynamics of the genome and to relate these features to its biological activities. The Network aims to generate integrated maps and comprehensive, quantitative models of nuclear organization in diverse cell types and conditions, including in single cells. Overall, we anticipate that these efforts will lead to new mechanistic insights into how the genome is organized, maintained, expressed, and replicated, in both normal and disease states.

Specifically, the 4DN Network will 1) develop, benchmark, validate, and standardize a wide array of technologies to probe the 4D nucleome; 2) integrate, analyze, and model datasets obtained with these technologies to obtain a comprehensive view of the 4DN; and 3) investigate the functional role of various structural features of chromatin organization in transcription, DNA replication and other nuclear processes. These three main components are illustrated in Figure 1.

To achieve these objectives, we have identified the following key steps. First, a set of common cell lines will be studied across the Network to enable direct cross-validation of data obtained with different methods (Table 1). Important criteria for selecting a set of common cell lines include a stable, haplotype-phased and normal karyotype, ease of growth, ease of genome editing, and suitability for (live-cell) imaging. Further, given that cell populations are characterized by significant cell-to-cell variation in their biological state (e.g. cell cycle stage), it will be important to employ clonal cell populations that can be synchronized, activated, induced or differentiated in a controlled manner.

Second, standards for data formats and data quality will be established so that data can be shared and compared among members of the consortium and the scientific community. This includes defining metrics for reproducibility across technical replicates (for technology development) and across biological replicates, and an objective assessment of the sensitivity, specificity, resolution and precision with which aspects of the 4D nucleome can be measured.

Third, computational and analytical tools will be developed to analyze individual datasets and to integrate, compare and cross-validate data obtained with different technologies. Importantly, they will allow for the integration of the diverse datasets necessary to build comprehensive models of the 4D nucleome.

Fourth, genetic, biochemical and biophysical approaches will be developed to measure and perturb the roles of DNA sequences and trans-acting factors (proteins, RNA) in the formation of local and global aspects of the 4D nucleome and their impact on transcription and other nuclear functions.

Fifth, a common vocabulary will be developed to describe nuclear features and biophysically derived principles guiding chromosome folding. This is important because currently different structural descriptions and interpretations have been put forward to describe features detected by different technologies, or even by the same methods. We need better descriptions of the underlying reality of structural features that make up the 4D nucleome, e.g. loops and domains, and develop a consistent terminology as these features are detected by different technologies. This can be achieved by integrated analysis of data obtained with the wide range of technologies employed and under development by the Network. This vocabulary, which will also include nomenclature for non-chromosomal nuclear features, will be of significant value to the community as a whole.

Finally, to enhance synergy between the different groups and prevent the unnecessary duplication of efforts, and to facilitate rapid dissemination of data to the larger scientific community, a shared database and repository will be established which includes all data, detailed protocols, engineered cell lines, and reagents used across the Network. The Network will build and maintain a public browser for all 4D Nucleome data to enable easy access to data and models

Structure of the 4D Nucleome Network

The 4DN Network encompasses several related efforts (http://www.4dnucleome.org/; Figure 2). First, six Centers make up the Nuclear Organization and Function Interdisciplinary Consortium (NOFIC). Each of these NOFIC Centers will develop genomic and imaging technologies, and implement computational models for understanding the 4D Nucleome. NOFIC centers will also work together with the other components of the Network to benchmark the performance of these experimental and computational tools, and to identify the most appropriate repertoire of methods to study the 4D Nucleome. These studies will be combined with extensive structural and functional validation of observations and models. Ultimately, the NOFIC aims to deliver integrated approaches that can be used towards generating a first draft of a model of the 4D nucleome.

Second, ongoing technology development will also be critical to this effort, and is addressed by the 4D Nucleome Network in three ways. (i) New genomic interaction technologies will be developed to study the 4D nucleome at the single cell level, for example, to detect higher-order clusters of interacting genomic loci, to probe the roles of RNA in chromatin architecture, and to engineer new chromatin interactions. (ii) Several groups will develop new imaging and labeling methods to visualize the genome at high resolution, in live cells as well as in tissues, and in relation to genome activity. (iii) New methods will be developed to probe the DNA, RNA and protein composition of subnuclear structures such as the nuclear envelope, the nucleolus and other nuclear bodies. These methods, including procedures to reversibly disrupt these structures, will allow mechanistic studies of the role of nuclear bodies/structures in chromatin organization and gene expression under various conditions of health and disease.

Third, a Data Coordination and Integration Center (DCIC; http://dcic.4dnucleome.org/) will store all data generated by the Network, and coordinate data analysis. The DCIC will maintain a web portal to share data, insights and models with the Network and the larger scientific community. A separate Organizational Hub will coordinate all communication across the Network and organize regular meetings for all participants of the Network.

Research Plans

Genomic technologies to reveal the 4D nucleome

Chromosome conformation capture (3C) technologies have been developed to examine long-range interactions across the genome [15,18]. Genome-wide 3C technologies, *e.g.* Hi-C, have revealed patterns of interaction that define genome structures at various resolutions, including topologically associating domains (TADs) [16,19,20]. TADs can be hundreds of kilobases in size, often containing several genes and multiple enhancers, at least some of which appear to interact by looping mechanisms. The ChIA-PET method provides finer resolution to detect detailed domain structures defined by chromatin architectural proteins such as CTCF and cohesin, as well as specific enhancer-promoter interactions associated with RNAPII and other TFs [6]. Furthermore, genome-wide mapping at base pair resolution to detect haplotype specific interactions are in progress, which will enable the connection of chromatin topology to the vast genetic information regarding complex traits and diseases [17,36,37]. The results generated by these and related methods have been tremendously exciting, and the Network will continue to develop, optimize, and integrate 3C-based technologies with the goal of further elucidating the different levels of chromatin interactions in the nucleus, from the local nucleosome organization of chromatin polymers to their global 3D organization in domains, chromosome territories and in relation to other nuclear structures (Box 1). These efforts will be complemented by orthogonal approaches including direct imaging technologies at both standard and super resolution, X-ray microscopy, and electron microscopy (below and Box 2).

The goal of producing large numbers of data sets of moderate resolution (~1-5 kb) with today's technology is being complemented by an ongoing effort to invent improved methods for interaction mapping aimed at delivering additional genome interaction information at higher resolution and with haplotype-resolved specificity, lower costs, and at the single cell level. Additional development will also be focused towards implementing genome-wide strategies for mapping RNA-DNA interactions (Box 1).

A major limitation of nearly all 3C methods is that they depend on a single cross-linker, formaldehyde, or on double cross-linking reagents that combine EGS (ethylene glycol bis-succinimidylsuccinate) followed by formaldehyde [38]. Because of the nature of formaldehyde, which is known to polymerize and crosslink molecules across a large range of distances, this approach complicates the interpretation of the observed chromatin contacts as it lacks precise distance information. The Network will explore novel bivalent photo-activated cross-linkers, which have specificity for DNA and are separated by linkers of defined length and flexibility. Other strategies under development entirely eliminate the use of cross-linkers. Additionally, novel genome-wide methods that enable exploration of higher-order (beyond pairwise) DNA contacts, e.g. genomic DNA regions that form hubs, at various resolutions are being developed. Finally, new methods are being developed that can assess three dimensional genome

organization in (thousands of) individual cells [39]. Clearly, a diversity of existing and new approaches will be necessary to resolve the precise nature of interactions and the physical distances separating the interacting DNAs as well as the cell-to-cell variation in these patterns. The resulting data will provide important information for subsequent modeling of the structural and functional nature of these interactions.

Imaging the 4D genome

4DN investigators will develop and integrate multi-modal imaging platforms that enable visualizing the dynamics, interactions, and structural organization of the nucleus at unprecedented temporal and spatial resolutions (Box 2). Each of these imaging modalities will have unique and complementary abilities for probing different aspects of genome organization both across the cell cycle and during differentiation. In particular, platforms enabling live cell imaging uniquely allow the dynamics of select chromatin regions and nuclear features to be studied in real time.

Both standard and high-throughput fluorescent in situ hybridization (FISH) using oligonucleotide probes or guide RNA-mediated recruitment of fluorescently labeled dCas9 ("CASFISH" [33]) in fixed cells will be exploited to image pairwise and multiplexed genomic interactions over different spatial distances in different cell types and states. In addition, CRISPR/dCas9 FISH in live cells [31,32,40] and other live cell imaging approaches (Box 2) will be employed to assay the dynamic behavior of particular chromatin regions and/or nuclear structures in real time. These imaging tools will play an important role in benchmarking, validating and complementing data obtained with genomic and proteomic mapping technologies.

New technologies will be developed to exploit chemical, nucleic acid, enzymatic and fluorescent probes to label DNA, RNA, and protein that occur in proximity of specific nuclear bodies. These technologies include HRP-labeled antibodies in TSA-SEQ (Box 1), APEX (ascorbate peroxidase; [41]) and Killer Red [42]. Many of these same chemicals and probes are compatible with more established imaging approaches that will also be used to visualize and validate these nuclear body proximity maps. Genome-editing technologies will also be applied to tag a subset of key genomic loci, loops, TADs and potentially newly discovered structures to help visualize these moieties and document their interactions with other nuclear regions in live cells.

Super resolution microscopy, single-molecule tracking techniques and multiplex fluorescent/chemical tags will be applied in living cells to determine the dynamic interactions, diffusion and motion of fluorescently labeled proteins, ncRNAs and genomic loci (Box 2). These live cell imaging approaches are expected to provide novel information regarding search mechanism, binding and residence time of both DNA and protein interactions and will also be relied on extensively to validate and complement genomic methods used by the Network. For example, these methods will be used to determine whether the different contact frequencies of loci in 3D genomic maps reflect their relative distances in real space and/or their interaction frequency and dynamics.

Both imaging and genomics groups will focus on particular TADs and loops to define high resolution solutions to the interaction domains of these structures with respect to both other chromatin domains and nuclear bodies using the same cell clones growing under defined conditions.

Soft X-ray Tomography (SXT) will be used to visualize the 3D organization of chromatin in nuclei of cells in the native state (cryo-immobilized). SXT will be used to directly measure chromatin compaction, e.g. in relation to sub-nuclear position, at different stages of cell differentiation and in different cell types. Correlated microscopy approaches will be used to augment ultrastructural data with molecular localization information. Cryogenic Fluorescence Tomography (CFT) will be used to precisely locate molecules in 3D reconstructions of intact, native state cells imaged using Soft X-ray Tomography.

Members of the Network will develop new Electron Microscopy (EM) technologies that enable the local and global structural organization of chromatin to be visualized as a continuum from nucleosome to Mbp scale in both interphase and mitotic cells. One such method, ChromEM, will be combined with new genetic tags and nanoparticle labeling technologies to develop the EM equivalent of 'multi-color' fluorescence.

The development of automated imaging analysis pipelines and data standards will be important to extract the maximum structural information possible from these datasets. Further development of software for analyzing, annotating, and archiving imaging data, together with implementation of new approaches for correlating imaging and genomics datasets are major goals of the 4DN Network (see below).

Nuclear bodies and non-chromatin structures

The nucleus consists of distinct nuclear structures as well as non-chromatin bodies, including the nuclear lamina, nuclear pores, nuclear speckles, and nucleoli [43,44]. Increasing evidence indicates that specific genomic regions associate with these structures, suggesting that these chromosomal associations may have a functional role in regulating genome function [8,45,46].

Major challenges today revolve around our poor understanding and characterization of both the structure and function of these nuclear structures and bodies. Future progress in understanding the relationship between nuclear structure and function demands new methodologies to study the proximity of different chromosomal loci to these different nuclear regions as well as to describe how chromosomal fibers traverse and/or move between these different nuclear regions. Similarly, we envision new technologies to interrogate the protein and RNA content of these different nuclear regions. Additionally, we need high-throughput approaches and imaging methods that query the correlations between genome localization near these different nuclear regions and changes in the functional outputs of these genomic regions. Likewise, we need new methods to dissect the mechanisms by which large genomic regions are targeted to specific nuclear

structures / bodies / compartments. Finally approaches to perturb these nuclear structures / bodies /compartments and/or to redirect chromosome regions to different nuclear regions will allow assessing the functional significance of nuclear compartmentalization.

Goals of the 4D Nucleome Network further include development of new mapping methodologies to measure the genome-wide molecular interaction frequency and cytological distance of chromosome loci to major nuclear compartments, including the nuclear lamina, nuclear pores, nuclear speckles, nucleoli, and pericentric heterochromatin (Box 1). Concurrently, new and improved technologies will be developed, including localized APEX (ascorbate peroxidase)-mediated protein biotinylation [47], fractionation by cryomilling, and RNA Antisense Purification (RAP) ([48-50]), to catalog and measure the protein and RNA components of these nuclear compartments, as well as both optogenetic and degron-based approaches to alter or disrupt sub-nuclear bodies and compartments. Functional mapping approaches based on Repli-Seq [9] and TRIP (thousands of reporters integrated in parallel; [51]) will provide genome-wide correlations of DNA replication timing and chromosome position effects on transcription and RNA processing that can be correlated with these structural maps (Box 1). New imaging approaches will be developed to correlate chromosome and nuclear compartment dynamics with changes in DNA replication timing, transcriptional activation, and other functional states (Box 2). Computational analyses of these genome mapping data from several cell types will be aimed at identifying possible *cis* and *trans* determinants of nuclear compartmentalization.

Modeling the 4D Nucleome

As described above, the 4DN network will develop an increasingly powerful repertoire of tools for generating high-throughput data that inform us about 3D/4D nuclear architecture. The major classes of methods yield data that are informative with respect to biochemically-derived contact probabilities of loci (e.g. 3C and its derivatives) or the spatial distances between loci (e.g. FISH and other imaging methods). Additional technologies yield data that are informative of other aspects of nuclear architecture (e.g. the relative positions and contents of nuclear structures in relation to one another as well as the folded genome).

In parallel and often in concert with the emergence of these experimental methods has been the development of computational approaches for modeling the spatial organization of the genome. There are at least two major computational approaches for global modeling of genome architecture on the basis of experimental data: *data-driven* and *de novo* approaches [34,35]. *Data-driven approaches* directly use experimental data to produce a single conformation or an ensemble of conformations that best match an experimentally observed set of contact probabilities or distances [52]. *De novo* modeling, on the other hand, produces ensembles of conformations that result from known or hypothesized physical or biological processes, and tests whether these ensembles are consistent with features of experimental contact frequency maps and imaging data.

Such *de novo* models can suggest specific molecular mechanisms and principles of chromosome organization, thus going far beyond the experimental data [35].

Although these approaches to modeling genome architecture have been useful for making sense of the observed data and relating it to principles of chromosome organization, it is clear that we are still far from a high-confidence, high-resolution map of genome architecture. Therefore, an overarching goal of the Network is to generate reliable and validated structural maps representing the 4D Nucleome for different biological states of the genome.

A number of challenges and opportunities with respect to computational modeling of genome architecture will be addressed. First, computational models of global genome architecture have typically relied on a single data type, e.g. Hi-C, which limits the power of such analyses because it is likely that Hi-C or ChIA-PET or any other one method does not capture all aspects of the 4D nucleome and does not sufficiently constrain models. The Network will apply a wide diversity of technologies, and each of these is likely to capture complementary aspects of genome organization, e.g. contacts by Hi-C and distances by imaging [53,54]. A major challenge is thus to adapt current modeling approaches so that data obtained with these methods (e.g. Hi-C and imaging data) can be systematically integrated and used to generate comprehensive structural and dynamic models of the 4D nucleome. Second, nearly all current genomic methods yield data that are derived from an ensemble of thousands to millions of cells, obscuring structural heterogeneity that exists among single cells. A number of groups within the Network are developing methods for generating global data from large numbers of single cells, which will present new computational challenges for integrating with current modeling approaches [39]. It is possible that some of these methods will yield functional data on these same single cells (i.e. Hi-C and RNA-seq, from each of many single cells), which would represent a significant opportunity for directly relating structure to function. Third, most current models ignore the fact that mammalian cells are diploid, i.e. they do not distinguish or separately model homologous chromosomes, which will be particularly important for modeling based on single-cell data. It is critical to develop experimental methods and computational pipelines that overcome this deficiency [17,36,37]; the haplotype-resolution of the genomes of the common cell types chosen by the consortium will aid in this goal. Fourth, contemporary approaches for modeling genome architecture are typically static rather than dynamic, reflecting the static nature of available Hi-C data and the majority of imaging data. As we are increasingly able to visualize (via direct imaging) or infer (via single cell or bulk Hi-C analyses of time series, live cell imaging) dynamic processes, e.g. differentiation, response to physiological stimuli, and cell cycle progression, it will be essential that these observations can be integrated into computational models. Fifth, new data and future models should help to connect genome architecture and other aspects of genome function, e.g. by suggesting molecular mechanisms of how transcription factor binding or epigenetic modifications can lead to formation of active/inactive chromatin compartments [55,56], etc. The inferred relationships between these features and genome architecture would serve as a means of generating testable predictions of sequence-structure-function relationships, i.e. how nuclear architecture relates to nuclear function.

Perturbation and manipulation of the 4D Nucleome to relate structure to function

A critical and overarching goal is to determine how genome structure and chromatin interactions at multiple scales modulate genome function in health and disease. However, despite significant progress, many of the mechanisms by which chromosome structures such as chromatin domains are formed and how they are related to local genome activity remain poorly understood.

Therefore, the 4DN Network will explore a range of experimental approaches to specifically manipulate and perturb different features of the 4DN. First, using CRISPR/Cas9 technologies and targeted DNA integration experiments, DNA elements involved in specific chromatin structures, e.g. domain boundaries or chromatin loops, can be altered, re-located or deleted [57,58]. Second, defined chromatin structures such as chromatin loops will be engineered de novo by targeting proteins that can (inducibly) dimerize with their partner looping proteins (e.g. [7]). Third, other CRISPR/Cas9 approaches will be used to target enzymes (e.g. histone modifying enzymes, structural proteins) or ncRNAs to specific sites in the genome. Fourth, several groups will perturb nuclear compartmentalization by developing methods for "rewiring" chromosome regions to different nuclear compartments, either by integrating specific DNA sequences that are capable of autonomous targeting the locus to different nuclear compartments or bodies or by tethering certain proteins to these chromosomal loci to accomplish similar re-positioning. Fifth, cell lines will be generated for conditional/temporal ablation of nuclear bodies or candidate chromosome architectural proteins (such as CTCF and cohesin) or RNAs. Sixth, additional methods will be developed to nucleate nuclear bodies at specific chromosomal loci. Finally, biophysical approaches will be developed to micro-mechanically perturb cell nuclei and chromosomes followed by direct imaging of specific loci [59]. Analysis of the effects of any of these perturbations on processes such as gene expression and DNA replication will then provide deeper mechanistic insights into the roles of chromosome structure and nuclear organization in regulating the genome.

Data sharing and standards

An important goal for the Network will be to develop practical guidelines for data formats, metadata (detailed descriptions of how the data was acquired), standards, quality control measures, and other key data-related issues. Another goal will then be to make this production-quality data rapidly and readily accessible, both within the Network and the entire scientific community. These efforts will be of particular importance for new technologies for which standards for sharing data and assessing data quality have not yet been established. Such standards will greatly enhance the usefulness of the datasets for the broader scientific community beyond those who generate the data.

For sequencing-based technologies, data format standards to represent sequences and alignments have long been present (e.g. fastq, bam/sam). However, common formats to represent three-dimensional interactions are yet to be developed. These formats need to account for large data sizes and the constraints imposed by different computer architectures. For Hi-C data, for example, the genome-wide contact probability map is an N^2 matrix, where N is the spatial resolution (for 10kb resolution, N=300,000), with most of the entries being empty; whereas for ChIA-PET data the detected chromatin contacts are between selected loci bound by a protein of interest, which greatly reduces the complexity of the interaction matrix. There are multiple ways to represent such sparse matrices, appropriate for different analysis and storage approaches. For imaging technologies, the situation is even more challenging, as the types of microscopes employed are highly variable, and the data formats and analysis tools are often manufacturer-dependent. Standards to unify data and metadata from different manufacturers, such as the Open Microscopy Environment (https://www.openmicroscopy.org/site), are under development. These standards also need to accommodate the rapid and ongoing developments in super-resolution microscopy technologies.

A related issue is to define a set of appropriate metadata fields and minimum metadata requirements such that sufficient and useful details are available to other investigators outside the Network. While it is acknowledged that not all information can be captured about an experiment, collecting pertinent information (without unnecessarily burdening the experimentalists) will increase the reproducibility of experiments and the likelihood that the data will be re-used by other investigators. Towards this end, the 4DN Consortium has established formal working groups, including the 4DN Data Analysis Group, Omics Standards Group and Imaging Data Standards Group, which will play critical roles in defining these 4DN standards and data analysis protocols.

Developing a set of measures for assessing data quality and determining appropriate thresholds will be important for ensuring high-quality 4DN data. An important measure of data quality and reliability of new technologies is the reproducibility of results between repeated experiments. Reproducibility can be assessed at multiple levels: broadly speaking, technical reproducibility measures how well a technique performs for the same starting material, whereas biological reproducibility should also capture all other variations including heterogeneity among samples. Data quality standards and thresholds for what constitutes 'good' data will change over time as technologies improve and analysis methods are refined. To accommodate this fluidity, the 4DN Analysis Group will compute and make available relevant quality control measures and provide recommendations on expected quality standard thresholds so that investigators can make decisions regarding the utility of specific datasets for addressing their specific questions.

Outlook

After determining the complete DNA sequence of the human genome, the variations in this sequence in individuals and across populations, and mapping most genes and

potential regulatory elements, we are now in a position that can be considered the third phase of the human genome project. In this phase, the spatial organization of the genome is elucidated and its functional implications revealed. This requires a wide array of technologies from the fields of imaging, genomics, genetic engineering, biophysics, computational biology and mathematical modeling. The 4D Nucleome Network, as presented here, provides a mechanism to address this uniquely interdisciplinary challenge. Further, the policy of openness and transparency both within the Network and with the broader scientific community, and the public sharing of all methods, data and models will ensure rapid dissemination of new knowledge, further enhancing the potential impact of the work. This will also require fostering collaborations and establishing connections to other related efforts around the world that are currently under development.

Together these highly integrated studies, both within the 4DN Network, and with partners around the world, promise to allow moving from a one-dimensional representation of the genome as a long DNA sequence to a spatially and dynamically organized three dimensional structure of the living genome inside cells. This structure will place genes and regulatory elements that can be far apart along the DNA, in a three-dimensional and dynamic context, revealing their functional relationships and providing a new understanding of how the genome is regulated.

Figure 1. The 4D Nucleome project. The project encompasses three components:
First, experimental mapping approaches are employed to measure a range of aspects
of the spatial organization of the genome including chromatin loops, domains, nuclear
bodies etc. Second, computational and modeling approaches are used to interpret
experimental observations and build (dynamic) spatial models of the nucleus. Third,
perturbation experiments, e.g. using CRISPR/cas9-mediated genome engineering, are
used for functional validation. In these studies chromatin structures are altered, e.g.
removing chromatin loops or inducible creating novel loops at defined positions. These
perturbation studies can be complemented with functional studies, e.g. analysis of gene
expression to assess the functional implications of chromatin folding. (Picture of cell
nucleus was provided by Hanhui Ma and Thoru Pederson).
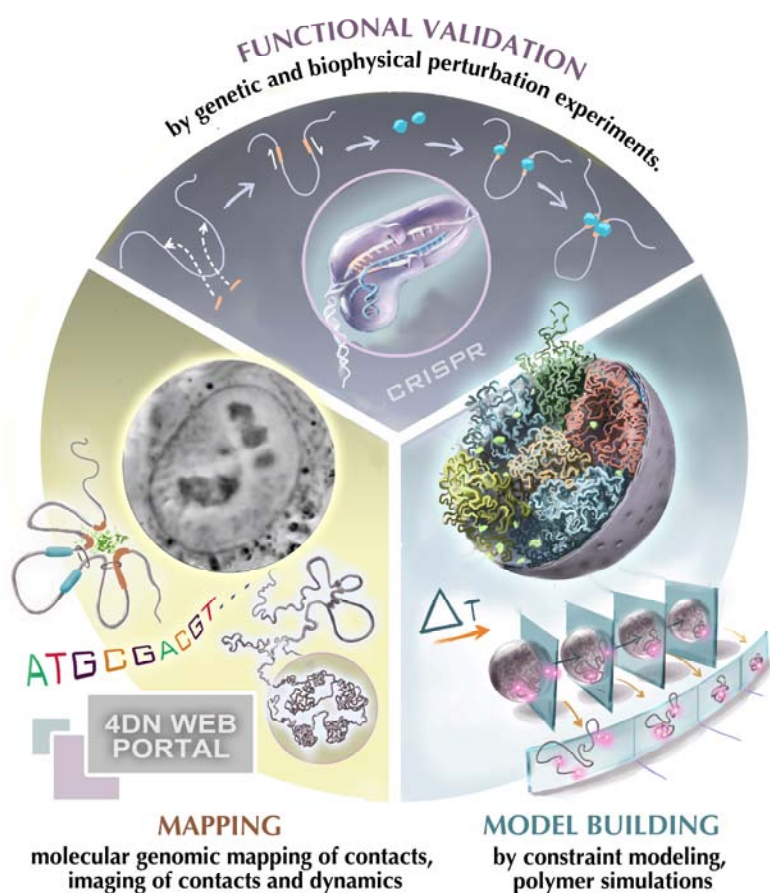
Figure 2. **Organization of the 4D Nucleome Network.** Six interdisciplinary centers make up the Nuclear Organization and Function Interdisciplinary Consortium. Three additional groups of funded projects focus on analysis of nuclear bodies and development of nucleomics and imaging tools. A Data Coordination and Integration Center collects all data for analysis and public sharing. The Organizational Hub coordinates all Network activities.
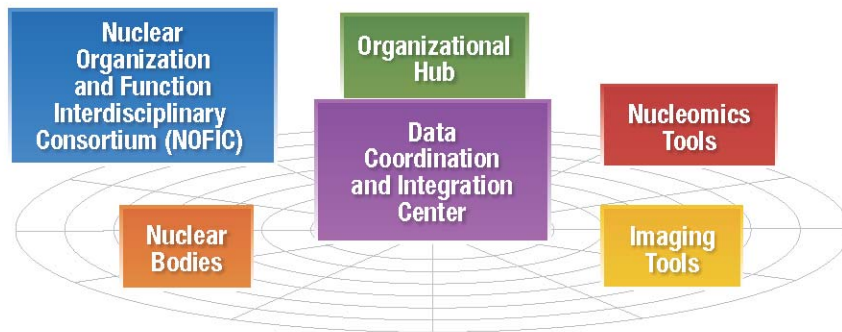
Table 1: Common cell lines used by the 4D Nucleome network. Tier 1 cells are used for all studies, while Tier 2 cell lines are used for specific projects.

Tier 1	Description	Availability
H1-ESC (WA01)	Human embryonic stem cells (male)	WiCell
hTert-HFF	hTert-immortalized human foreskin fibroblasts (male)	Dekker lab
IMR90	Primary breast fibroblasts (female)	ATCC
GM12878	EBV-immortalized B-lymphocytes (female)	Coriell
F121-9	Mouse Cast-129 F1 hybrid embryonic stem cells (female)	Gilbert lab

Tier 2	Description	Availability
K562	Chronic myelogenous leukemia (female)	ATCC
hTert-RPE	hTert-immortalized retina pigmented epithelium (female)	ATCC
HEK293	Human embryonic kidney (likely female)	ATCC
U-2 OS	Osteosarcoma cells (female)	ATCC
HAP1	Haploid derivative of chronic myelogenous leukemia (CML) cell line KBM-7 (male)	Horizon Genomics
WTC-11	iPSC (male)	TBD
F123	Mouse Cast-129 F1 hybrid embryonic stem cells (male)	TBD
H9-ESC (WA09	Human embryonic stem cells (female)	TBD

Box 1: Genomic technologies currently in use or in development in the 4DN network

Assay Name	Assay Abr.	Key features	Reference	Related Protocols
Chromosome Conformation Capture		All methods use proximity ligation to measure interacting genomic loci		69-71
	3C	Interactions between specific genomic loci, including genome-wide studies	15,60	
	4C	Genome-wide interactions of a specific genomic locus	61	72,73
	5C	Many loci against many loci	62	
	Hi-C	Genome wide map of all interactions in the nucleus	16	
	Single Cell Hi-C	Hi-C variant that enables mapping contacts within single cells	39	
	Combinatorial single-cell Hi-C	Single cell Hi-C variant using split-pool barcoding to map single cells	http://biorxiv.org/content/early/2016/07/23/065052	
	In situ **Hi-C**	Hi-C variant that performs digestion and ligation in intact nuclei	17	
	DNase Hi-C	Hi-C variant that digests chromatin using DNase	63	
	Micro-C	Hi-C variant that digests chromatin using micorcoccal nuclease	64	74
	Capture Hi-C	Hi-C variant that incorporates selection of targeted genomic loci	65,66	
	TCC	A variant of Hi-C using bead-coupling of complexes	67	
	Distance Hi-C	Variant of Hi-C using photo-activated crosslinkers to measure the distance between interacting DNA regions.		
	COLA	A variant of Hi-C using frequent restriction cutters to map >2 simultaneous DNA interactions	68	
Chromatin Interaction Analysis by paired-end-tag sequencing	ChIA-PET	Genome-wide map of interactions co-occurring with a specific protein.	6,38	
Genome Architecture Mapping	GAM	A cryosectioning method to get interacting DNA regions in a ligation-independent manner. Can achieve higher-order interactions.	Under development	
Split-pool		A ligation-independent method using a split-	Under	

barcoding of RNA and DNA		and-pool strategy to barcode interacting RNA and DNA in the nucleus that enables genome-wide mapping of higher-order contacts.	development	
RNA interaction with chromatin by paired-end-tag sequencing	RICh-PET	Genome-wide mapping for all ncRNA-chromatin interactions.	Under development	
Chromatin IP aided RICh-PET	ChIP RICh-PET	Genome-wide mapping for 3-way interactions involving ncRNA, DNA loci, and protein factors.	Under development	
DNA Adenine Methyltrans-ferase Identification	DamID and Single-Cell DamID	Genome-wide mapping of molecular contact frequency of DNA locus to a Dam methylase fusion protein: will be used to measure DNA proximity to different nuclear compartments	75,8,76; under development	77,78
Tyramide Signal Amplification-Seq	TSA-Seq	Genome-wide mapping of estimated mean cytological distance (in microns) of DNA locus to a nuclear compartment; will be used to map cytological distances to several nuclear compartments	Under development	
Replication Labeling-Seq	Repli-Seq	Genome-wide mapping of timing of DNA replication	79,80	81
Thousands of Reporters Integrated in Parallel	TRIP	Genome-wide mapping of chromosome position effects on transcription and post-transcriptional processing	51	82
RNA Antisense Purification	RAP	Systematic mapping of DNA regions, RNA species, and proteins that are in proximity to a specific noncoding RNA of interest in the nucleus.	48-50	83,84

Box 2: Imaging technologies currently in use or in development in the 4DN network

Application	Method	Label/Stain/Dye	Feature	References
Visualizing DNA and RNA sequences	3D DNA and RNA FISH	Fluorescently labeled DNA and RNA oligo probes	Fixed cells	
	HIPMap	High throughput FISH	High throughput FISH and automated microscopy in fixed cells	85
	LacO and TetO labeled genomic loci	Lac-O and Tet-O repeats inserted into genome, visualized by fluorescently labeled lac and tet repressor proteins	Can be used in live and fixed stable cell-lines	86
	Cas-FISH and Cas live cell imaging	Fluorescently labeled dCas9 and guide RNAs to label genomic loci	Live and fixed cells- guide RNAs against repetitive targets and/or collections of guide RNAs that cover unique sequences	32,40,33
	CRISPRainbow	gRNA contains stem loop that is bound by fluorescently labeled coat protein	Live cell imaging of repetitive DNA targets and/or unique sequences using collections of gRNAs	31,87
	MERFISH	Multiple detection of up to 1000 mRNA species		
	TALE-FISH	Fluorescent TALE fusions that label specific genomic loci with naturally occurring repeats	Live and fixed cells	88,89
	MS2, PP7, mSpinach, aptamers	Fluorescent fusions that bind and label tagged RNA sequences	Live and fixed cells	
Labels and tags	Fluorescent	GFP, mCHERRY, Killer-RED, Dendra Suntag, SH2 Quantum dots	Imaging live and fixed cells-diffraction limited	
	Photo-activated and Photo-switchable	PA-GFP/MCherry/mKATE TAG-RFP, Dronpa, Alexa-488/568/647 PS-CFP2, Eos, atto488, Cy5 etc	super resolution stochastic imaging, single-molecule tracking, live and fixed cells	
	Peroxidases Correlated light and EM	HRP, APEX, APEX2 miniSOG ChromEM ALEXA 633 Fluoronanogold Time-STAMP-YFP-MiniSOG	For labeling in electron microscopy, sometimes combined with correlative light and electron microscopy (CLEM)	

	Clickable tags	Janelia Fluor dyes: 646, 549 etc Snap and Halo tags	Genetically encoded tags for live and fixed cell imaging	90
High Resolution Ultrastructure and 3D organization	Transmission EM	Heavy metal stains- Osmium tetroxide, Uranyl acetate, lead. Colloidal goal and nanogold labels. CLEM inorganic and organic probes and tags	2D transverse projection through 70-250 nm sections	
	Serial Block Face Scanning EM (SBEM)		Reconstruction and scanning EM images of serial sections (70 nm) through entire cell/tissue volume	
	Mult-tilt EM Tomography		High resolution ultrastructure of macromolecules in situ through large 3D volumes (250-400 nm thick sections) of the nucleus	
	Serial section EMT (ssEMT)		Reconstruction of larger 3D volumes (>500nm thick) by alignment of serial EMT datasets	
	X-Ray Tomography	Quantitative linear absorption coefficient (LAC) generated contrast that reflects the bioorganic composition of unlabeled DNA, RNA and proteins	Mesoscale resolution (20-50 nm) of nuclear structure in intact, unprocessed cells	91
	Correlated SIM and X-Ray Tomography	Fluorescently labeled probes	Visualizing the structure and interactions of selectively labeled chromatin and protein interactions in intact, unprocessed whole cells	92
	Multi-Color EM (EELS and EMT)	Orthogonal correlated light and EM probes and DAB conjugates to visualize chromatin and proteins/ RNA in the nucleus	Visualizing the ultrastructure and interactions of selectively labeled chromatin and protein interactions simultaneously	93
	ChromEMT	ChromEM labeling of chromatin and 8-tilt EMT	Enables the local ultrastructure and global 3D organization of chromatin as a continuum at nucleosome resolutions and megabase scales in single cells	Under development

Visualizing spatial and dynamic 3D nuclear organization	Wide-field, Confocal/Multiphoton 3D SIM PALM STORM STED Adaptive Optics tcPALM Lattice Light Sheet (LLS) LLS-PAINT Single-molecule tracking in live cells FCS FRAP FRET FLIP		Diffraction limited (>250 nm) Super-resolution imaging of nuclear organization in fixed cells (ca. 20 nm). 3D resolution in PALM/STORM Nuclear organization in live cells 3D dynamics and nuclear organization Dynamics Allows measurements of binding and search dynamics (residence time, search time, search mechanism)	94 30 95,96 97-99

References

1   ENCODE-Project-Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

2   Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).

3   Stunnenberg, H., International Human Epigenome Consortium & Hirst, M. The international human epigenome consortium: a blueprint for scientific collaboration and discovery. *Cell* **167**, 1145-1149 (2016).

4   Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563 (2005).

5   Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Mol Cell* **10**, 1453-1465 (2002).

6   Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627 (2015).

7   Deng, W. *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849-860 (2014).

8   Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951 (2008).

9   Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402-405 (2014).

10   Maurano, M. T. *et al.* Sytematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).

11   Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301 (2001).

12   Markaki, Y. *et al.* Functional nuclear organization of transcription and DNA replication: a topographical marriage between chromatin domains and the interchromatin compartment. *Cold Spring Harb Symp Quant Biol.* **75**, 475-492 (2010).

13   Bickmore, W. A. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet.* **14** (2013).

14   Politz, J. C., Scalzo, D. & Groudine, M. The redundancy of the mammalian heterochromatic compartment. *Current opinion in genetics & development* **37**, 1-8, doi:10.1016/j.gde.2015.10.007 (2016).

15   Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306-1311 (2002).

16   Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

17   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).

18   Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357-1382 (2016).

19   Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

20   Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).

21   Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038-2049 (2016).

22   Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A.* **112**, E6456-6465 (2015).

23   de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell.* **60**, 676-684 (2015).

24   Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297-1309 (2015).

25   Dekker, J. & Mirny, L. A. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110-1121 (2016).

26   Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol.* **17**, 756-770 (2016).

27   Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal* **91**, 4258-4272, doi:10.1529/biophysj.106.091116 (2006).

28   Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645, doi:10.1126/science.1127344 (2006).

29   Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793-795, doi:10.1038/nmeth929 (2006).

30   Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664-667 (2013).

31   Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature biotechnology* **34**, 528-530, doi:10.1038/nbt.3526 (2016).

32   Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491, doi:10.1016/j.cell.2013.12.001 (2013).

33   Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc Natl Acad Sci U S A* **112**, 11870-11875, doi:10.1073/pnas.1515692112 (2015).

34   Marti-Renom, M. A. & Mirny, L. A. Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization. *PLoS Comput Biol* **7**, e1002125 (2011).

35   Imakaev, M. V., Fudenberg, G. & Mirny, L. A. Modeling chromosomes: Beyond pretty pictures. *FEBS letters* **589**, 3031-3036, doi:10.1016/j.febslet.2015.09.004 (2015).

36   Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).

37   Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579, doi:10.1038/nature18589 (2016).

38   Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64 (2009).

39   Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).

40   Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* **112**, 3002-3007, doi:10.1073/pnas.1420024112 (2015).

41   Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature biotechnology* **30**, 1143-1148, doi:10.1038/nbt.2375 (2012).

42   Bulina, M. E. *et al.* A genetically encoded photosensitizer. *Nature biotechnology* **24**, 95-99, doi:10.1038/nbt1175 (2006).

43   Dundr, M. Nuclear bodies: multifunctional companions of the genome. *Current opinion in cell biology* **24**, 415-422, doi:10.1016/j.ceb.2012.03.010 (2012).

44   Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends in genetics : TIG* **27**, 295-306, doi:10.1016/j.tig.2011.05.006 (2011).

45   Németh, A. *et al.* Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).

46   van Koningsbruggen, S. *et al.* High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell.* **21**, 3735-3748 (2010).

47   Lee, S. Y. *et al.* APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell reports* **15**, 1837-1847 (2016).

48   Engreitz, J. M. *et al.* RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188-199, doi:10.1016/j.cell.2014.08.018 (2014).

49   McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236, doi:10.1038/nature14443 (2015).

50   Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

51   Akhtar, W. *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**, 914-927, doi:10.1016/j.cell.2013.07.018 (2013).

52   Serra, F. *et al.* Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS letters* **589**, 2987-2995, doi:10.1016/j.febslet.2015.05.012 (2015).

53   Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome biology* **17**, 215, doi:10.1186/s13059-016-1081-2 (2016).

54   Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *bioRxiv* **doi:10.1101/081448** (2016).

55   Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* **42**, 9553-9561, doi:10.1093/nar/gku698 (2014).

56   Pombo, A. & Nicodemi, M. Physical mechanisms behind the large scale features of chromatin organization. *Transcription* **5**, e28447, doi:10.4161/trns.28447 (2014).

57   Xiong, X., Chen, M., Lim, W. A., Zhao, D. & Qi, L. S. CRISPR/Cas9 for Human Genome Engineering and Disease Research. *Annual review of genomics and*

human genetics **17**, 131-154, doi:10.1146/annurev-genom-083115-022258 (2016).

58   Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem* **85**, 227-264, doi:10.1146/annurev-biochem-060815-014607 (2016).

59   Poirier, M. G. & Marko, J. F. Micromechanical studies of mitotic chromosomes. *Current topics in developmental biology* **55**, 75-141 (2003).

60   Rodley, C. D., Bertels, F., Jones, B. & O'Sullivan, J. M. Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet Biol.* **46**, 879-886 (2009).

61   Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348-1354 (2006).

62   Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements. *Genome Res.* **16**, 1299-1309 (2006).

63   Ma, W. *et al.* Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods.* **doi: 10.1038/nmeth.3205** (2014).

64   Hsieh, T. S. *et al.* Mapping nucleosome resolution chromosome folding in yeast by Micro-C. . *Cell* **162**, 108-119 (2015).

65   Dryden, N. H. *et al.* Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* **24**, 1854-1868 (2014).

66   Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205-212 (2014).

67   Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol.* **30**, 90-98 (2011).

68   Darrow, E. M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**, E4504-4512, doi:10.1073/pnas.1609643113 (2016).

69   Hagège, H. *et al.* Quantitative analysis of Chromosome Conformation Capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722-1733 (2007).

70   Louwers, M., Splinter, E., van Driel, R., de Laat, W. & Stam, M. Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nature protocols* **4**, 1216-1229, doi:10.1038/nprot.2009.113 (2009).

71   Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using chromosome conformation capture. . *Methods* **in press** (2012).

72   Splinter, E., de Wit, E., van de Werken, H. J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221-230, doi:10.1016/j.ymeth.2012.04.009 (2012).

73   van de Werken, H. J. *et al.* 4C technology: protocols and data analysis. *Methods in enzymology* **513**, 89-112, doi:10.1016/B978-0-12-391938-0.00004-5 (2012).

74   Hsieh, T. 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, doi:10.1038/nmeth.4025 (2016).

75   van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol.* **18**, 424-428 (2000).

76   Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-147, doi:10.1016/j.cell.2015.08.040 (2015).

77   Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nature protocols* **2**, 1467-1478, doi:10.1038/nprot.2007.148 (2007).

78   Pindyurin, A. V., Pagie, L., Kozhevnikova, E. N., van Arensbergen, J. & van Steensel, B. Inducible DamID systems for genomic mapping of chromatin proteins in Drosophila. *Nucleic Acids Res* **44**, 5646-5657, doi:10.1093/nar/gkw176 (2016).

79   Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).

80   Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A.* **107**, 139-144 (2010).

81   Ryba, T., Battaglia, D., Pope, B. D., Hiratani, I. & Gilbert, D. M. Genome-scale analysis of replication timing: from bench to bioinformatics. *Nature protocols* **6**, 870-895, doi:10.1038/nprot.2011.328 (2011).

82   Akhtar, W. *et al.* Using TRIP for genome-wide position effect analysis in cultured cells. *Nature protocols* **9**, 1255-1281, doi:10.1038/nprot.2014.072 (2014).

83   Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).

84   Simon, M. D. *et al.* The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* **108**, 20497-20502, doi:10.1073/pnas.1113536108 (2011).

85   Shachar, S., Voss, T. C., Pegoraro, G., Sciascia, N. & Misteli, T. Identification of Gene Positioning Factors Using High-Throughput Imaging Mapping. *Cell* **162**, 911-923, doi:10.1016/j.cell.2015.07.035 (2015).

86   Robinett, C. C. *et al.* In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol.* **135**, 1685-1700 (1996).

87   Ma, H. *et al.* CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *The Journal of cell biology* **214**, 529-537, doi:10.1083/jcb.201604115 (2016).

88   Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc Natl Acad Sci U S A* **110**, 21048-21053, doi:10.1073/pnas.1319097110 (2013).

89   Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M. E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nature structural & molecular biology* **20**, 1321-1324, doi:10.1038/nsmb.2680 (2013).

90   Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* **12**, 244-250, 243 p following 250, doi:10.1038/nmeth.3256 (2015).

91   Le Gros, M. A. *et al.* Soft X-Ray Tomography Reveals Gradual Chromatin Compaction and Reorganization during Neurogenesis In Vivo. *Cell reports* **17**, 2125-2136, doi:10.1016/j.celrep.2016.10.060 (2016).

92   Smith, E. A. *et al.* Quantitatively imaging chromosomes by correlated cryo-fluorescence and soft x-ray tomographies. *Biophysical journal* **107**, 1988-1996, doi:10.1016/j.bpj.2014.09.011 (2014).

93   Adams, S. R. *et al.* Multicolor Electron Microscopy for Simultaneous Visualization of Multiple Molecular Species. *Cell chemical biology* **23**, 1417-1427, doi:10.1016/j.chembiol.2016.10.006 (2016).

94   Izeddin, I. *et al.* PSF shaping using adaptive optics for three-dimensional single-molecule super-resolution imaging and tracking. *Optics express* **20**, 4957-4967, doi:10.1364/OE.20.004957 (2012).

95   Chen, B. C. *et al.* Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998, doi:10.1126/science.1257998 (2014).

96   Legant, W. R. *et al.* High-density three-dimensional localization microscopy across large volumes. *Nat Methods* **13**, 359-365, doi:10.1038/nmeth.3797 (2016).

97   Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285, doi:10.1016/j.cell.2014.01.062 (2014).

98   Izeddin, I. *et al.* Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **3**, doi:10.7554/eLife.02230 (2014).

99   Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF abd Cohesin regulate chromatin loop stability with distinct dynamics. *bioRxiv* **doi: https://doi.org/10.1101/093476** (2016).