nature genetics

PERSPECTIVE

https://doi.org/10.1038/s41588-018-0236-3

Challenges and guidelines toward 4D nucleome data and model standards

Marc A. Marti-Renom[1,2,3*], Genevieve Almouzni[4], Wendy A. Bickmore[5], Kerstin Bystricky[6], Giacomo Cavalli[7], Peter Fraser[8,9], Susan M. Gasser[10,11], Luca Giorgetti[10], Edith Heard[12], Mario Nicodemi[13,14], Marcelo Nollmann[15], Modesto Orozco[16,17], Ana Pombo[14,18,19] and Maria-Elena Torres-Padilla[20,21]

Due to recent advances in experimental and theoretical approaches, the dynamic three-dimensional organization (3D) of the nucleus has become a very active area of research in life sciences. We now understand that the linear genome is folded in ways that may modulate how genes are expressed during the basic functioning of cells. Importantly, it is now possible to build 3D models of how the genome folds within the nucleus and changes over time (4D). Because genome folding influences its function, this opens exciting new possibilities to broaden our understanding of the mechanisms that determine cell fate. However, the rapid evolution of methods and the increasing complexity of data can result in ambiguity and reproducibility challenges, which may hamper the progress of this field. Here, we describe such challenges ahead and provide guidelines to think about strategies for shared standardized validation of experimental 4D nucleome data sets and models.

The genome represents much more than a passive library of genetic information. It functions as an information-retrieval device instructed to dynamically change conformation to expose the genetic information required for a particular cell type under a particular cellular situation. It is now thought that distinct changes in the three-dimensional (3D) arrangement of our (epi)genomes occur in association with development, physiological responses, aging, response to diet, environmental stress and/or learning. Just as a map of the world is more than a list of places and street names, the nuclear genome is more than a string of letters and can be described as a complex choreography of proteins and nucleic acids that interact differentially over time. How these complex DNA-based functions are achieved within the constrained nuclear space, and how the chromatin context of any given gene is regulated, remained elusive until recently.

New technologies are paving the way for exploration of the structure and dynamics of chromatin[1,2], and future developments integrating different approaches offer further promises. Indeed, we now face the great challenge of taking the linear genome sequence provided by the Human Genome Project[3,4], decorated with the valuable annotations provided by the ENCODE, Roadmap and FANTOM projects, among others[5], and creating an integrated 4D understanding of the complexity of the cell nucleus. Given the remarkable technologies and data sets now available, it is time to launch a concerted effort toward characterizing the dynamic organization of the genome and the rules that govern determination and maintenance of cell types (see, for example, recent complementary initiatives in US, NIH-4DNucleome[6]; Europe, EU-4DNucleome and LifeTime; and Japan, Japan-4DNucleome). We envisage a complete 3D atlas in time (4D) of nuclei within the thousands of cell types that form an organism.

The 4D nucleome is a rapidly evolving field that has been exponentially growing since the 1980s, just as the protein structure field exploded in the 1960s once the first structures of proteins emerged. The establishment of the Protein Data Bank (PDB)[7] and its associated structural format was instrumental for the sharing of information, allowing structural data to revolutionize biochemistry, protein engineering approaches and drug design. However, as huge amounts of structural information accumulated, it became clear that to maximize the utility and reliability of the structural models, standardized and validated pipelines for data processing were needed, and the massive storage of standardized data sets were required in addition to mere 3D coordinates. First the X-ray crystallography community[8,9], and later other groups[10,11], decided to address this challenge, and the PDB has been evolving toward a standardized, validated metadata repository. The 4D nucleome community is at a turning point and must now address a similar challenge.

[1]Gene Regulation, Stem Cells and Cancer Program, CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [3]ICREA, Barcelona, Spain. [4]Institut Curie, PSL Research University, CNRS UMR3664, Paris, France. [5]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine at the University of Edinburgh, Edinburgh, UK. [6]Laboratoire de Biologie Molécualire Eucaryote (LBME), Centre de Biologie Intégrative (CBI), University of Toulouse, CNRS, bat. IBCG, Toulouse, France. [7]Institute of Human Genetics, UMR 9002 of the CNRS and the University of Montpellier, Montpellier, France. [8]Nuclear Dynamics Programme, The Babraham Institute, Cambridge, UK. [9]Department of Biological Science, Florida State University, Tallahassee, Florida, USA. [10]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [11]University of Basel, Basel, Switzerland. [12]Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris, France. [13]Dipartimento di Fisica "E. Pancini", Università di Napoli "Federico II", INFN Sezione di Napoli, Naples, Italy. [14]Berlin Institute of Health, MDC-Berlin, Berlin, Germany. [15]Centre de Biochimie Structurale, CNRS UMR5048, INSERM U1054, Universite de Montpellier, Montpellier, France. [16]Institute for Research in Biomedicine Barcelona (IRB), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [17]Departament de Bioquimica i Biologia Molecular, Universitat de Barcelona, Barcelona, Spain. [18]Epigenetic Regulation and Chromatin Architecture, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany. [19]Institute for Biology, Humboldt-Universitat zu Berlin, Berlin, Germany. [20]Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum Munchen, München, Germany. [21]Faculty of Biology, Ludwig-Maximilians Universitat, München, Germany. *e-mail: martirenom@cnag.crg.eu

Uncorrected Proofs

Assessment of genome and epigenome structural and dynamic data from molecular genomics, imaging and computational modeling requires agreement on a series of validation pipelines that the community should adopt, as well as data sharing strategies and an accepted format for depositing the data. This is a nontrivial issue given the rapid evolution of technologies, the variety of methods used and the intrinsically multidimensional nature of the problem. Moreover, proper translation of the data into biological insight with subsequent functional validations represents a major challenge. Next, we briefly review the state of the art of each of the standardization and validation approaches to study and interpret nuclear organization.

Molecular genomics

The 4D nucleome community has developed and implemented genomic technologies that allow an integrated investigation of gene expression, epigenetic marks, nucleosome localization and genome interactions, both for cell populations and at the single-cell level. These technologies rely on the use of high-throughput experimental approaches and next-generation sequencing (NGS) to characterize genome organization and chromatin status at the molecular level. Information retrieved from sequencing provides data of protein occupancy at specific genomic regions, nucleosome organization, DNA methylation and 3D chromatin associations, which enable multiresolution information on chromatin structure to be deduced and modeled.

For example, nucleosome positioning techniques are based on sequencing of chromatin after DNA digestion that treats nucleosome-bound compared to unbound (linker) DNA differentially. Several technologies have been developed that differ mainly in the way DNA is digested or how DNA–protein interactions are identified[12,13]. Promoter and enhancer accessibility data are key for understanding regulation of gene expression, and assay for transposase-accessible chromatin using sequencing (ATAC)-seq has emerged as a tool to probe this aspect[14]. All these approaches present challenges that need to be addressed to validate and standardize the resulting data sets, especially for single-cell experiments in which the data are sparse and cannot be replicated. For example, the data can be noisy, biased by sequence preferences of processing enzymes, cross-linking reagents, antibody specificity, and standard sequence biases of NGS. Moreover, nucleosomes are sliding along the sequence, which means they do not have fixed positions, but probability distributions are observed, which require very deep sequencing to be reliable. Importantly, nucleosome positioning data sets almost invariably represent an average map, which has to be considered in light of cell-to-cell variability, something that requires complex deconvolution of the detected signals into unique nucleosome families. Recently, single-cell and single-molecule mapping variants of some of these technologies provided an exciting perspective while raising the challenge of gathering epigenomic maps from intrinsically scattered data[15–17].

Although these approaches provide a steady state picture of the epigenome, an emerging class of time-resolved methods should ultimately enable its dynamic characterization at multiple time scales. Indeed, histone dynamics have been studied genome-wide on a time scale of minutes[18–20]. An emerging microfluidics-based methodology also allows tracking of site-specific protein–DNA contact kinetics on a time scale of seconds by measuring cross-linking kinetics[21]. Other approaches use fusion constructs to characterize protein residence time on DNA, exploiting a time course of MNase digestion[22] or the anchor-away system[23]. Although the diversity of experimental systems provides great promise, comparing these methods has proven difficult. The building of integrated models of the 4D nucleome should tackle these limitations and solve the problem of translating linear nucleosome distributions into time-dependent 3D arrangements consistent with time-resolved imaging data.

Chromosome conformation capture techniques (3C)[24] come in various flavors[25]. These approaches are limited by a number of potential issues, including the degree of cross-linker fixation, nuclear permeabilization, as well as solubilization, digestion and ligation efficiencies. Other more specific issues such as oligonucleotide design exist for 5C and capture Hi-C. Data processing and bioinformatic analyses of "C" data are complex, and approaches for quality control and normalization remain an open challenge. This is particularly important given the blooming of experimental variations on the main technology, each carrying different advantages, limitations and potential sources for bias. Common standards do not yet exist, which is becoming a major challenge for the ability to compare different data sets to obtain meaningful conclusions. We need to perform extensive benchmarking of the experimental and computational analyses to reach a consensus on standards[26]. Furthermore, validation with independent technologies is needed. Direct comparison with data sets from orthogonal approaches, such as those from genome architecture mapping (GAM)[27], and imaging offers great promise toward defining gold standards[28–31]. Though this should mature, newer approaches will continue to emerge. For example, in GAM[27], C-Walks[32] or SPRITE[33], more than two chromatin contacts can be captured and analyzed. This calls for setting up dedicated analytical tools and pipelines and further testing with complementary technologies, such as high-throughput microscopy.

Light and electron microscopy

DNA and RNA FISH have long been used to evaluate the location and activity of genomic loci and the position of whole chromosomes in situ at the single-cell level. The limitations in spatial resolution, throughput, and genomic coverage have been major challenges. However, the advent of 3D high-throughput (deep-imaging) and super-resolution imaging (nanoscopy) technologies coupled with novel DNA/RNA labeling strategies are now enabling the visualization of genomic domains, individual genes and single transcripts in 3D. Novel approaches to tag chromosomal domains include multiplexing probes by oligopaints or antibody labeling in fixed cells[30,34–36]. Well-known technologies for live imaging involve the use of stem-loop structures in RNAs and fluorescent viral proteins that recognize these tags (MS2/MCP or PP7)[37,38], as well as bacterial operator arrays that can be visualized by fluorescently tagged repressor proteins and other "fluorescent repressor operator systems" (FROS)[39–41]. New labeling techniques take advantage of protein oligomerization such as Suntag[42] or ANCHOR[43]. The advent of CRISPR-inactive Cas9-bound guide RNAs for fluorescently tagging DNA[44,45] or transcription activator-like effectors (TALE)[46,47] is changing the landscape of visualization options when fused to fluorescent proteins to enable the visualization of naturally occurring repetitive sequences. Each of these systems is likely to have its specific limitations and may potentially interfere with chromatin biology in a context-dependent manner as reported for lacO FROS[48] or CRISPR-based tagging systems[49]. However, several reports recapitulate the live-cell imaging results (TetO-tagged loci in embryonic stem cells) in fixed cells[50,51], suggesting that these approaches are not generally disruptive of genome architecture and function. SNAP- or CLIP-tag technologies[52,53] for labeling proteins, fluorescent antibody fragments[54] and development of new, brighter dyes offer seemingly unlimited possibilities to probe nanoscale structures of chromatin. As these methods rely on fusion proteins (with potential hindrance to their function) or antibodies (with potential specificity issues), integrating results from multiple approaches will be critical. Photoactivable or SNAP-tag-based approaches can also be used to measure protein dynamics in the nucleus[55], together with fluorescence recovery after photobleaching (FRAP), fluorescence correlation spectroscopy (FCS) and single-molecule tracking methods. Analytical methods to integrate data from these approaches are a subject of intense research[31,35]. In the past 10 years,

Uncorrected Proofs

super-resolution imaging of single molecules has paved the way for even an more detailed analysis of chromatin structural variations within the nucleus[56,57]. In particular, it is now possible to beat the diffraction limit by a full order of magnitude using stochastic optical reconstruction microscopy (STORM), photoactivated localization microscopy (PALM) and structured illumination microscopy (SIM), which can reach 20-nm lateral resolution.

An experimental challenge for single-molecule imaging in live nuclei is the need to capture the rapid 3D motion of nuclear factors (diffusion coefficient ~10 μm²/s). The recently developed multifocal microscope (MFM) enables the parallel acquisition of up to twenty-five focal planes[58,59]. One can thus image cellular volumes over an axial depth of ~5 μm, comparable to the size of eukaryotic nuclei, with acquisition rates of up to hundreds of times faster than conventional methods[60,61]. Standard fluorescence signals can now also be captured in thick samples with light-sheet microscopy, which, together with aberration-corrected multifocus microscopes, will enable time-lapse imaging that is many fold faster and has much less bleaching than conventional spinning-disk or wide-field fluorescence microscopy[58]. In fixed cells, chromatin fibers can be visualized by chromatin EM tomography[62]. Combined light microscopy–electron microscopy (CLEM) provides locus-specific labeling by fluorescence and either conventional EM or serial block face scanning electron microscopy ultrastructural information of the entire nucleus, cell or tissue. 3D EM can now also be correlated with time-lapse fluorescence imaging in living cells[63]. Finally, SBF-EM[64] enables 3D reconstruction of cellular structures at nanometer resolution.

Altogether, it is now clear that chromatin structure is becoming accessible to the microscope on all levels of resolution, as are long-range contacts through the 3C technologies. Indeed, light imaging has enabled partial validation of "C" data[35,65], as the distributions of distances between loci can be measured and chromatin domains or TADs mapped using large or multiple probes[30,31,35,65]. Furthermore, FISH or live-cell tagging technologies can be combined with immunofluorescence[66] or fluorescent fusion proteins for a variety of nuclear bodies and structures[51,67–69] to gain insight into the chromatin state or nuclear compartment in which a locus resides. The most obvious advantage of these imaging approaches is that the position and status of a genomic locus can be measured in vivo and at the single-cell level in a cell population and in tissue sections, preserving cell–cell interactions. The disadvantages still remain the number of loci that can be measured at any one time, the need to genetically modify, transfect or otherwise treat the cells to generate fluorescent signals and the reliance on in vitro models of cell culture. The controls and standards for these techniques remain very different among laboratories. In microscopy, not only the experimental conditions but also the instrument itself and the image-analysis tools provide a huge number of variables. Standard pipelines for quantification of signal adapted to the noise (pixels), segmentation procedures and explicit distinction between 2D and 3D image analysis are, with some exceptions, lacking in the field. Finally, an effort must be made to deploy image-analysis codes in an open web format.

A final and very real challenge for imaging is performing experiments that are physiologically relevant; that is, finding conditions that minimize phototoxicity and potential artifacts of fixation. As the biological relevance of any observed event is primordial, it must be tracked over multiple conditions, and the statistical analysis of many individual event recordings must be considered integral to any microscopic approach. In addition to the need for high-resolution and rapid time-lapse imaging, we require means of storing, processing and analyzing the huge imaging data sets, as well as analytical tools that extract physical principles from the geometry and movement of chromatin. Only then can modeling of particle and fiber dynamics be applied to properly interpret the results of moving chromatin loci and integrate them with 'fixed' cell imaging or population-based molecular analyses.

3D/4D modeling

Theoretical approaches have become key for investigating the complexity of high-throughput 4D nucleome data. Two main approaches have been used over the last decades to model chromatin[70]. First, physical modeling, which has its roots in atomistic simulation methods, has been used to identify plausible spatial arrangements of the chromatin fiber, consistent with the laws of polymer physics. For example, these approaches have been implemented to interpret the decay of interaction frequencies with the genomic distance[71,72], the formation of domains of active and inactive chromatin[73], the formation of domains by loop extrusion[74], and epigenetic features such as chromatin types[75], chromosome territories[76–78] and co-expression data[79], among others. Furthermore, models have been directly compared to experimentally derived interaction data sets to characterize potential molecular mechanisms underlying chromosome folding[27,74,80–83]. Second, the so-called restraint-based modeling has more recently been used to represent experimental observations from cell populations and single cells as sets of spatial restraints to fold the genome in 3D. The main experimental data driving these models come from 3C-based experiments, including single-cell Hi-C[84–86], as well as imaging[87].

Independently of the approach used for 3D/4D modeling, validating the accuracy of the resulting structural models is important and challenging. The difficulties arise not only from the diversity of chromatin arrangements in the cell population and the limited information about the native configuration/organization of the genome, but also from the lack of standards for sharing models. Currently, there is not a proposed standard for storing the coordinates of the resulting models, for linking the models with the experimental data used for its derivation, or for putting them in the context of previously known genomic information. More dramatically, no protocol exists to bridge the different levels of resolution on genome architecture and allow the final user to navigate from the base pair to the nucleosome array up to the global chromatin structure. Nowadays, each software uses its own formats, which are not easy to share, and key data to reproduce the computational experiment are not stored, making reproducibility a big challenge. Only recently have a significant number of models of genome organization at different levels of resolution been published[65,88–94]. However, those are not centrally deposited, do not share standards and have only been partially validated. It is now time to identify guidelines for validating, annotating, and depositing 3D/4D genome models at all resolutions generated with software that can be easily shared and models that can be reproduced, mimicking efforts made by other communities. Initial steps toward validation and assessment of the resulting structural models are being taken[95], including efforts such as those implemented in the virtual research environment of the Multiscale Complex Genomics Consortium (MuG), and this work needs to be continued and generalized in the future.

Experimental systems and functional validation assays

The ultimate validation of the data and models will come from predictions and their experimental testing by perturbing the biological systems of interest[96]. With the recently extended toolbox and the rise of biophysics and computational biology, the characterization of 4D genome structure and function is booming. Though technology has been an important driver for this progress, answering critical questions in the field such as how much of the structure dictates function and vice versa requires the selection of the appropriate biological systems and approaches to dissect the role of putative regulatory components. This issue is not novel, and the ENCODE project[97] has proposed a series of cell lines and tissues that could be used for these studies, which are subdivided into tiers 1, 2 and 3 in decreasing

Uncorrected Proofs

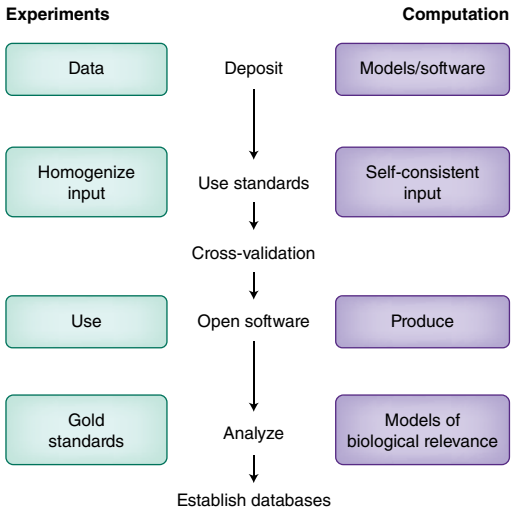

Fig. 1 | Toward FAIR data, unified standards and RICH visualizations in 4D nucleome research. Experiments and computation need to follow good practice guidelines for depositing, using, validating and analyzing standardized data sets.

Table 1 | Challenges in producing, analyzing, storing and disseminating experimental data and models for the 4D nucleome field

Experimental data	3D–4D modeling methods	Deposited models
■•Standard compliance. ■•Accessible and traceable. ■•Incorporate metadata (e.g., experimental set-ups). ■•Should be self-consistent. ■•Synthetic data for validation mimicking controlled experiments are needed for validation.	■•Standard compliance. ■•Clear definition of the basic physical assumptions. ■•Benchmarked. ■•Reliability metrics ■•Stable, traceable and accessible. ■•Flexibility to adapt to different experimental set-ups. ■•Cross-validate results with data not used in the refinement. ■•Compare results with random models.	■•Standard compliance. ■•Accessible and traceable. ■•Sustainable extension model to avoid missed information. ■•Incorporate metadata on the experimental and modeling parts. ■•Should be ready for RICH visualizations and integrated in multiresolution browsers for functional annotations.

order of priority. Tier 1 cell lines have been used extensively to analyze chromatin contacts[98,99], but, to understand functional aspects of chromosome architecture dynamics, it is important to analyze the temporal order of events using nontransformed cells that can be differentiated or stimulated. For this reason, models such as human H1 or mouse embryonic stem (ES) cells that allow differentiation into several lineages[100–103], cell stimulation by hormone treatment[104] or genomic perturbations of key structural proteins[105,106] can provide invaluable information. The collection of dense time series will be important for characterizing regulatory and stochastic differences over time. Likewise, analyses of multiple interrelated cell types along well-characterized differentiation paths, such as in the hematopoietic system[107], are likely to be instrumental in the elaboration of predictive models. The analysis of carefully controlled cell differentiation systems should be complemented by in vivo work from sorted cells[102], and particularly in the case of humans, a valuable alternative is the use of organoids obtained from differentiation of ES or iPS cells.

Furthermore, to go beyond correlations, it is important to manipulate the genome and uncouple effects on chromosome structure from effects on gene expression and DNA replication or repair. Multiple approaches are being developed toward this goal. In addition to knockdown and CRISPR-mediated gene knockout approaches, TALE- and CRISPR-based technology can be used to tether regulatory factors to study the effect on their target sites. This is true both for activation[108] and for repression[109]. Furthermore, recent developments allow multiplexing of these gene regulatory approaches[110]. Another strategy consists in tethering DNA domains to nuclear landmarks such as the lamina, which allows testing whether 'geographical' changes in gene position affect chromosome architecture and function[111]. TALE-based approaches can also be used to alter chromatin condensation without affecting transcription to study the effects on nuclear positioning of the cognate loci[112,113]. Moreover, specific mutations can be induced at individual sequence elements to study whether changes at a given position in the genome can induce long-range effects elsewhere[114,115]. Finally, new predictive models, such as PRISMR, are now able to predict the effect of structural variants in the topology of the genome[96]. Further developments of this rich toolbox should allow the performance of surgical genome technology experiments that should help researchers characterize the role of chromatin components in a quantitative way and tease apart correlations from causative roles.

Toward FAIR data, unified standards and RICH visualizations in 4D nucleome research

The quantity and quality of information that is being generated to explore and understand the organization, packaging and functions of the genome has grown tremendously in recent years, but many of these data cannot be fully exploited owing to the lack of standardization of procedures, lack of detailed information on various aspects of the setup and lack of data formats and visualization. Likewise, many current models cannot be used or compared among each other. To fully exploit the information that is being generated, both at the experimental and at the theoretical level, all these data should be made findable, accessible, interoperable and re-usable (FAIR)[116]. Here, we would like to propose good practice guidelines to ensure that data follow FAIR principles and to facilitate the implementation of common standards, both on the experimental and the modeling side (Fig. 1 and Table 1).

Concerning experimental approaches, we propose the following recommendations: (i) Data deposition. Deposit data in public repositories, providing rich and detailed metadata describing the materials, biological samples, experimental conditions and protocols. For imaging, the development of public repositories is still in discussion and in the testing phase, but it is highly recommended that primary images with appropriate metadata be stored and maintained until public repositories become available to the community. (ii) Standards. Use standardized, benchmarked experimental protocols for sample preparation and analysis. If the approach involves establishing new strategies, accompany new data with a standard data set to allow comparison with previous work. (iii) Homogenize. Reduce cellular heterogeneity by maximizing cell-type purity, reducing cell numbers studies and comparing cells in the same cell cycle stages. For single-cell studies, provide one replicate of bulk cells and sufficient numbers of single cells to allow merging of libraries to compare single-cell results with bulk population experiments. (iv) Validate data orthogonally. For instance, Hi-C data may be validated by using DNA FISH or by

Uncorrected Proof

other genomic approaches such as GAM. Likewise, in vivo imaging using GFP fusion derivatives can be validated by FISH and/or IF with appropriate antibodies. Finally, different super-resolution microscopy technologies should be compared to cross-validate a portion of the results of any given series of new experiments. In addition, data predictions made by chromatin contact data may be validated by other methods, such as DamID, or by testing interactions of chromatin associated proteins with techniques such as FRET or BiFC. These validations can be used to set up or improve modeling approaches. (v) Use open software. Analyze the data by using publicly available pipelines with software code available in full. When developing new analytical approaches, both for genomics and for imaging technologies, benchmark new software when possible and make the code publicly and freely available to the community. (vi) Set with gold-standards. Standard samples could be agreed upon by the community so that groups adopting a new technique or developing novel methods can have a benchmark to validate and compare their new approaches. This will be key to ensure reproducibility and for validating the real advantages of new methods compared to established ones. (vii) Establish resources databases. The field would considerably profit from the establishment of resources where genomic and microscopy data can be deposited, which would encourage cross validation from other scientists of primary results, allow proper benchmarking of new analysis methods or pipelines and encourage the use of machine learning or other emerging technologies to combine data from different sources to unveil novel mechanisms.

Likewise, researchers could consider the following points when developing tools for modeling[117]: (i) Comply with standards. Develop software that are properly benchmarked and provide measures of reliability. (ii) Self-consistent input. Be certain that the data used for modeling is self-consistent and does not result a significant portion of contradictory models. (iii) Produce models of biological relevance. The resulting models shall reflect the native dynamics of the genome and provide predictable and experimentally testable hypotheses. (iv) Capture variability. Ensure that the models agree with the assumptions of one or multiple states observed by experiments. (v) Cross-validate. Compare the models with experimental data sets not used during modeling (e.g., imaging data compared to 3C-based data). (vi) Analyze the models. Generate models for RICH visualization[118] to analyze them for additional nonrandom patterns that were not evident from used experimental data.

The time is ripe for the 4D nucleome community to discuss standards for the validation, deposition and analysis of data, as well as the resulting models needed for studying the spatiotemporal organization of the nucleus. Such criteria and standards could be inspired by the previous work that the structural biology community has carried out over the past decades for storing, disseminating, and visualizing data sets and models of proteins, nucleic acids and complex assemblies at different resolutions. To define these standards, we would like to propose that current initiatives in the 4D nucleome field collect recommendations by data producers and users, develop a consensus on validation protocols and identify software applications to perform such validation tasks. International experts in the 4D nucleome from experimental and computational fields, as well as visualization and data archiving, should meet to address a series of open questions that would bring this emerging field significantly closer to the desired standards. Finally, 4D nucleome research is young and booming. It can be predicted that data types and formats, as well as modeling procedures and capacities will evolve significantly in the coming years. It will therefore be important to evolve data and analytical standards correspondingly and ensure that old data, and not only the conclusions derived from them, will be 'lifted over to the newly defined standards to continue to be usable in the future, such that 4D nucleome knowledge will continue to increase without losing its history.

URLs. NIH-4DNucleome https://commonfund.nih.gov/4dnucleome/index; EU-4DNucleome, http://www.4dnucleome.eu; Japan-4DNucleome: https://doi.org/10.1080/19491034.2015.1022703; MuG, https://www.multiscalegenomics.eu; LifeTime, https://lifetime-fetflagship.eu

Received: 19 July 2017; Accepted: 19 July 2018;

References

1. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
2. Van Bortle, K. & Corces, V. G. Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.* **28**, 163–187 (2012).
3. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
4. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Bernstein, B. E. et al. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
6. Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
7. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
8. Read, R. J. et al. A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412 (2011).
9. Westbrook, J., Feng, Z., Burkhardt, K. & Berman, H. M. Validation of protein structures for protein data bank. *Methods Enzymol.* **374**, 370–385 (2003).
10. Montelione, G. T. et al. Recommendations of the wwPDB NMR Validation Task Force. *Structure* **21**, 1563–1570 (2013).
11. Sali, A. et al. Outcome of the First wwPDB hybrid/integrative methods task force workshop. *Structure* **23**, 1156–1167 (2015).
12. Teif, V. B. Nucleosome positioning: resources and tools online. *Brief. Bioinform.* **17**, 745–757 (2016).
13. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
14. Buffry, A. D., Mendes, C. C. & McGregor, A. P. The functionality and evolution of eukaryotic transcriptional enhancers. *Adv. Genet.* **96**, 143–206 (2016).
15. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
16. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
17. Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
18. Dion, M. F. et al. Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**, 1405–1408 (2007).
19. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
20. Deaton, A. M. et al. Enhancer regions show high histone H3.3 turnover that changes during differentiation. *eLife* **5**, e15316 (2016).
21. Poorey, K. et al. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342**, 369–372 (2013).
22. Zentner, G. E., Kasinathan, S., Xin, B., Rohs, R. & Henikoff, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* **6**, 8733 (2015).
23. Grimaldi, Y., Ferrari, P. & Strubin, M. Independent RNA polymerase II preinitiation complex dynamics and nucleosome turnover at promoter sites in vivo. *Genome Res.* **24**, 117–124 (2014).
24. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
25. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
26. Forcato, M. et al. Comparison of computational methods for Hi-C data analysis. *Nat. Methods* **14**, 679–685 (2017).
27. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
28. Williamson, I. et al. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **28**, 2778–2791 (2014).
29. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 215 (2016).
30. Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).

Uncorrected Proofs

31. Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* **14**, 673–678 (2017).
32. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
33. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).
34. Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
35. Cattoni, D. I. et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* **8**, 1753 (2017).
36. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
37. Fenrich, K. K. et al. Long-term in vivo imaging of normal and pathological mouse spinal cord with subcellular resolution using implanted glass windows. *J. Physiol. (Lond.)* **590**, 3665–3675 (2012).
38. Bertrand, E. et al. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
39. Straight, A. F., Belmont, A. S., Robinett, C. C. & Murray, A. W. GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol.* **6**, 1599–1608 (1996).
40. 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).
41. Lassadi, I., Kamgoué, A., Goiffon, I., Tanguy-le-Gac, N. & Bystricky, K. Differential chromosome conformations as hallmarks of cellular identity revealed by mathematical polymer modeling. *PLOS Comput. Biol.* **11**, e1004306 (2015).
42. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
43. Saad, H. et al. DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet.* **10**, e1004187 (2014).
44. Ma, H. et al. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. USA* **112**, 3002–3007 (2015).
45. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
46. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl. Acad. Sci. USA* **110**, 21048–21053 (2013).
47. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M. E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20**, 1321–1324 (2013).
48. Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C. & Taddei, A. Tight protein-DNA interactions favor gene silencing. *Genes Dev.* **25**, 1365–1370 (2011).
49. Lenstra, T. L., Coulon, A., Chow, C. C. & Larson, D. R. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Mol. Cell* **60**, 597–610 (2015).
50. Masui, O. et al. Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell* **145**, 447–458 (2011).
51. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439–445 (2002).
52. Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat. Methods* **10**, 119–121 (2013).
53. Chao, J. A., Yoon, Y. J. & Singer, R. H. Imaging translation in single cells using fluorescent microscopy. *Cold Spring Harb. Perspect. Biol.* **4**, a012310 (2012).
54. Stasevich, T. J. et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
55. Ray-Gallet, D. et al. Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell* **44**, 928–941 (2011).
56. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145–1158 (2015).
57. Matsuda, A. et al. Condensed mitotic chromosome structure at nanometer resolution using PALM and EGFP- histones. *PLoS One* **5**, e12768 (2010).
58. Abrahamsson, S. et al. Fast multicolor 3D imaging using aberration-corrected multifocus microscopy. *Nat. Methods* **10**, 60–63 (2013).
59. Abrahamsson, S. et al. MultiFocus polarization microscope (MF-PolScope) for 3D polarization imaging of up to 25 focal planes simultaneously. *Opt. Express* **23**, 7734–7754 (2015).
60. Hajj, B., El Beheiry, M. & Dahan, M. PSF engineering in multifocus microscopy for increased depth volumetric imaging. *Biomed. Opt. Express* **7**, 726–731 (2016).
61. Oudjedi, L. et al. Astigmatic multifocus microscopy enables deep 3D super-resolved imaging. *Biomed. Opt. Express* **7**, 2163–2173 (2016).
62. Ou, H. D. et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
63. Wolff, G., Hagen, C., Grünewald, K. & Kaufmann, R. Towards correlative super-resolution fluorescence and electron cryo-microscopy. *Biol. Cell* **108**, 245–258 (2016).
64. Titze, B. & Genoud, C. Volume scanning electron microscopy for imaging biological ultrastructure. *Biol. Cell* **108**, 307–323 (2016).
65. Giorgetti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
66. Gotta, M. et al. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *J. Cell Biol.* **134**, 1349–1363 (1996).
67. Pollex, T., Piolot, T. & Heard, E. Live-cell imaging combined with immunofluorescence, RNA, or DNA FISH to study the nuclear dynamics and expression of the X-inactivation center. *Methods Mol. Biol.* **1042**, 13–31 (2013).
68. Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M. & Gasser, S. M. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J. Cell Biol.* **168**, 375–387 (2005).
69. Heun, P., Laroche, T., Shimada, K., Furrer, P. & Gasser, S. M. Chromosome dynamics in the yeast interphase nucleus. *Science* **294**, 2181–2186 (2001).
70. Marti-Renom, M. A. & Mirny, L. A. Bridging the resolution gap in structural modeling of 3D genome organization. *PLOS Comput. Biol.* **7**, e1002125 (2011).
71. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
72. Barbieri, M. et al. A model of the large-scale organization of chromatin. *Biochem. Soc. Trans.* **41**, 508–512 (2013).
73. Barbieri, M. et al. Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017).
74. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Reports* **15**, 2038–2049 (2016).
75. 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 (2014).
76. Emanuel, M., Radja, N. H., Henriksson, A. & Schiessel, H. The physics behind the larger scale organization of DNA in eukaryotes. *Phys. Biol.* **6**, 025008 (2009).
77. Hahnfeldt, P., Hearst, J. E., Brenner, D. J., Sachs, R. K. & Hlatky, L. R. Polymer models for interphase chromosomes. *Proc. Natl. Acad. Sci. USA* **90**, 7854–7858 (1993).
78. Münkel, C. & Langowski, J. Chromosome structure predicted by a polymer model. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **57**, 5888 (1998).
79. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. & Micheletti, C. Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput. Biol.* **9**, e1003019 (2013).
80. 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. USA* **112**, E6456–E6465 (2015).
81. Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016).
82. Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. & Marenduzzo, D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503–3512 (2016).
83. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
84. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
85. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
86. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
87. Serra, F. et al. Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett.* **589**(20 Pt A), 2987–2995 (2015).
88. Jhunjhunwala, S. et al. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**, 265–279 (2008).
89. Ferraiuolo, M. A. et al. The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* **38**, 7472–7484 (2010).

Uncorrected Proofs

90. Duan, Z. et al. A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
91. Umbarger, M. A. et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* **44**, 252–264 (2011).
92. Baù, D. et al. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).
93. 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).
94. Trussart, M. et al. Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat. Commun.* **8**, 14665 (2017).
95. Trussart, M. et al. Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. *Nucleic Acids Res.* **43**, 3465–3477 (2015).
96. Bianco, S. et al. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, 662–667 (2018).
97. Gerstein, M. B. et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
98. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
99. Heidari, N. et al. Genome-wide map of regulatory interactions in the human genome. *Genome Res.* **24**, 1905–1917 (2014).
100. Gaspard, N. et al. An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* **455**, 351–357 (2008).
101. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
102. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
103. Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249 (2018).
104. Le Dily, F. et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28**, 2151–2162 (2014).
105. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
106. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
107. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
108. Hu, J. et al. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* **42**, 4375–4390 (2014).
109. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
110. Yan, Q. et al. Multiplex CRISPR/Cas9-based genome engineering enhanced by Drosha-mediated sgRNA-shRNA structure. *Sci. Rep.* **6**, 38970 (2016).
111. Wijchers, P. J. et al. Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol. Cell* **61**, 461–473 (2016).
112. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
113. Therizols, P. et al. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346**, 1238–1242 (2014).
114. Bantignies, F. et al. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* **144**, 214–226 (2011).
115. Deng, W. & Blobel, G. A. Detecting long-range enhancer-promoter interactions by quantitative chromosome conformation capture. *Methods Mol. Biol.* **1468**, 51–62 (2017).
116. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
117. Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77**, 443–477 (2008).
118. Goodstadt, M. & Marti-Renom, M. A. Challenges for visualizing three-dimensional data in genomic browsers. *FEBS Lett.* **591**, 2505–2519 (2017).

Acknowledgements

We thank the community of researchers around the EU 4DNucleome Initiative for their continuous support.

Author contributions

All authors wrote and approved the manuscript. Q11

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to M.A.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Uncorrected Proofs