

# THE STRUCTURE AND FUNCTION OF CHROMATIN AND CHROMOSOMES

WILLIAM STAFFORD NOBLE<sup>1,2</sup>, C. ANTHONY BLAU<sup>1,3</sup>, JOB DEKKER<sup>4</sup>, ZHI-JUN DUAN<sup>3</sup> AND YI MAO<sup>1</sup>

<sup>1</sup>*Department of Genome Sciences, University of Washington*

<sup>2</sup>*Department of Computer Science and Engineering, University of Washington*

<sup>3</sup>*Department of Hematology, University of Washington*

<sup>4</sup>*Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School*

## 1. Background

One of the striking features of the eukaryotic nucleus is that chromosomes adopt preferred conformations that vary across different tissues and developmental stages. Many lines of evidence suggest the functional significance of genome architecture, including the existence of chromosomes territories [1], the correlation between radial positioning of chromosomes with gene density, transcriptional activity, replication timing and chromosome size [2-5], the tissue-specific spatial configurations of chromosomes relative to one another [6, 7], the co-localization of disparate DNA elements into functionally defined aggregates or “factories” [9-13], and the predisposition to human disease, including cancer by predisposition to acquisition of karyotypic defects [14-17]. Collectively, these various lines of evidence indicate an intimate relationship between genome structure and function. These findings also emphasize the need to characterize both local and global chromosome structure to understand the underlying regulatory mechanisms of various genome functions.

In broad terms, three types of molecular biology tools are currently available to characterize chromosome structure.

1. Microscopy-based imaging, including live cell imaging [18], DNA FISH [19], and tomography [20] are established but relatively low throughput methods.
2. DNA adenine methyltransferase identification (DamID) [21] can be used to characterize genome structures associated with nuclear landmarks or other proteins. This method requires the use of genetically engineered cell lines.
3. Finally, chromatin conformation capture (3C) [24] and related methods, including 4C, 5C, ChIA-PET [22, 23] and Hi-C [35], are relatively simple and have been widely adopted for studies of long-range chromatin interactions.

3C and its derivatives are now the most commonly used tools for characterizing chromosome structure [24-34]. During the past two years, three 3C-based methods have been developed to identify chromatin interactions genome-wide, revealing the folding principles of the human genome [35] and enabling the construction of 3D models of the genomes of haploid budding yeast [36] and fission yeast [37].

The regulation of eukaryotic gene expression depends on transcription factors and cofactors acting on functional DNA elements such as promoters, enhancers, silencers and insulators, which are situated in a three-dimensional space. While systematic efforts such as those of the NIH

ENCODE Consortium have led to high throughput strategies for rapidly annotating functional elements in the human genome, how these elements interact in three dimensions is not well understood. The focus of this workshop is the development and application of computational techniques that link genome structure to function.

## 2. Workshop Contributions

The workshop includes two invited talks and five talks that were selected on the basis of submitted abstracts.

- **Leonid A. Mirny** is a Professor in the Division of Health Sciences and Technology and Department of Physics at the Massachusetts Institute of Technology. His invited lecture, “Higher-order chromatin architecture: bridging physics and biology”, will describe computational simulations of human chromatin based on Hi-C data, which support the hypothesis that, on a megabase scale, human chromatin adopts a so-called *fractal globule* configuration. The resulting model can be used to test mechanisms responsible for chromatin structure formation during the cell cycle, to investigate cell-to-cell variability in chromatin structure, and to establish a possible link between chromosomal abnormalities in cancer and high-order chromatin structure.
- **Yijun Ruan** is a Senior Group Leader and the Associate Director of Genomic Technologies at the Genome Institute of Singapore. His invited lecture, “Chromatin interaction and transcription regulation,” will describe the successful use of Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) to map long-range chromatin interactions associated with CTCF and RNA polymerase II in human cells. These analyses suggest a central role for CTCF-mediated looping in the general landscape of chromosomal organization. The results also show that many genes are transcribed cooperatively in a process heavily influenced by chromatin organization.
- Damian Wójtowicz, Fedor Kouzine, Arito Yamane, Wolfgang Resch, Craig J. Benham, Rafael Casellas, David Levens and Teresa M. Przytycka. “Genome-wide mapping and computational analysis of non-B DNA structures *in vivo*.” This talk describes a new experimental method, called ssDNA-seq, that identifies occurrences of single-stranded DNA (ssDNA) across a genome. Applied to mouse B cells and human Raji cells, the assay indicates an enrichment of ssDNA near transcription start sites, suggesting a role for ssDNA in gene regulation. The data also provides insights into existing computational methods for predicting non-canonical local DNA structure.
- Zackary Scholl, Jianling Zhong and Alexander Hartemink. “Mapping transcription factories in *Saccharomyces cerevisiae*.” In this project, the authors integrate published yeast data sets related to genome architecture [36], transcription localization [38, 39] and polymerase II occupancy [40] to show statistical evidence for the presence of transcription factories.
- Thomas Asbury, Matt Mitman, William Arndt, Jijun Tang and Jim Zheng. “Modeling, Analyzing and Visualizing Human Genome in 3D.” This talk describes an ongoing effort

to develop next-generation software for interactively visualizing eukaryotic genomes in three dimensions. Using simulated data, the authors demonstrate the ability to visualize the 3D human genome in real time and at scales ranging from 1bp resolution to the entire genome.

- Geet Duggal, Hao Wang, Michelle Girvan, Sridhar Hannenhalli and Carl Kingsford. “Graph-theoretic approaches for testing embeddability and spatial functional enrichment in chromosome conformation data.” In this project, the authors apply graph rigidity theory to detect regions of the yeast genome in which observed genome architecture data [36] is sufficient to identify a rigid embedding in three-dimensional space. This procedure yields a single, large component comprising the entire genome, with sub-components corresponding to each chromosome. The authors also propose a statistical test for co-localized sets of genes.

### **3. Outlook**

Similar to the rest of the genomics field, progress in elucidating genome structure and relating structure to biological function is driven largely by technological innovations. As such, future directions for the field are somewhat difficult to guess. Nonetheless, given current technologies, we can identify a few key areas in which concerted effort by computational biologists will likely lead to significant insights.

#### ***3.1. Statistical Confidence Estimation***

Large-scale genome structure assays such as Hi-C generate as their primary output a huge collection of sequence reads. Relatively robust, user-friendly tools exist for assigning these reads in an efficient fashion to a reference genome; however, such tools are currently lacking for the subsequent steps of aggregating reads with respect to, for example, restriction enzyme fragments and of assigning statistical confidence estimates to observed interactions. Existing statistical confidence estimation methods use straightforward empirical or analytic null models, and these models have not been extensively validated [35-37].

#### ***3.2. Inference of 3D Structure***

Given information about the frequency of pairwise interaction between pairs of DNA loci, as produced by medium- or high-throughput methods such as 4C, 5C or Hi-C, a natural next step is to attempt to infer properties of the 3D conformation of the DNA in the nucleus. Several successful attempts have been made along these lines, including a relatively straightforward, direct optimization approach [36, 37], as well as more sophisticated statistical sampling methods [41]. The latter approach is clearly preferable, because it has the potential to model cell-to-cell variability in 3D structure. Clearly, any 3D inference algorithm must take into account physical constraints imposed by the spatial organization of the nucleus itself as well as the physical properties of the DNA polymer. Furthermore, any solution to the 3D inference problem will likely need to be capable of integrating multiple types of data, including 3C-based measurements as well as information from FISH, DamID or ChIA-PET. Finally, a particular challenge for 3D

inference is appropriately handling multiploid genomes, and the fact that large populations of cells are analyzed, in which the observed count of interactions between any given pair of loci in the cell and in the cell population is actually a mixture of counts from all possible pairs of homologous copies of those loci. For more details on this general problem, see the recent excellent review by Marti-Renom and Mirny [42].

### **3.3. Visualization of Genomic Structure**

Widely used genome browsers, such as the Santa Cruz Genome Browser [43], the Ensembl Genome Browser [44] or JBrowse [45], are designed to work with data that is arrayed linearly along the genome. As such, representing pairwise interaction data within these frameworks is difficult. The Genome3D viewer [46], which is described in one of our submitted talks, is an excellent first step in this direction, but a fully functional platform that integrates the wide variety of available 1D and 3D data requires further effort.

### **3.4. Statistical Methods for Co-localization Analysis**

As pointed out above, one primary hypothesis driving the current interest in genome structure is that structure is causally linked to genome function in some non-trivial fashion. Accordingly, there is a growing demand for statistical methods that can bridge the gap between our current knowledge of gene function, as represented, for example, by Gene Ontology annotations or various functional genomics assays, and empirical observations of genome architecture. For example, recent evidence suggests that co-regulated yeast genes tend to be co-localized [47]. In this workshop, the contributed talk by Scholl *et al.* represents a novel approach to this type of problem. Ultimately, the field is aiming toward developing a physical model that suggests how genome form affects function.

## **References**

1. T. Cremer and M. Cremer. Chromosome territories. *Cold Spring Harb Perspect Biol.* **2**(3), a003889 (2010). PMID: 2829961.
2. S. Boyle, S. Gilchrist, J. M. Bridger, N. L. Mahy, J. A. Ellis and W. A. Bickmore. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet.* **10**(3), 211-219 (2001).
3. J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague and W. A. Bickmore. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol.* **145**(6), 1119-1131 (1999). PMID: 2133153.
4. F. Grasser, M. Neusser, H. Fiegler, T. Thormeyer, M. Cremer, N. P. Carter, et al. Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J Cell Sci.* **121**(Pt 11), 1876-1886 (2008). PMID: 2687722.
5. T. Misteli. Beyond the sequence: cellular organization of genome function. *Cell.* **128**(4), 787-800 (2007).
6. L. A. Parada, P. G. McQueen and T. Misteli. Tissue-specific spatial organization of genomes. *Genome Biol.* **5**(7), R44 (2004). PMID: 463291.

7. I. Solovei, M. Kreysing, C. Lanctot, S. Kosem, L. Peichl, T. Cremer, et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell*. **137**(2), 356-368 (2009).
8. H. Tanabe, S. Muller, M. Neusser, J. von Hase, E. Calcagno, M. Cremer, et al. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A*. **99**(7), 4424-4429 (2002). PMID: 123664.
9. M. Thompson, R. A. Haeusler, P. D. Good and D. R. Engelke. Nucleolar clustering of dispersed tRNA genes. *Science*. **302**(5649), 1399-1401 (2003).
10. G. L. Zhou, L. Xin, W. Song, L. J. Di, G. Liu, X. S. Wu, et al. Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol Cell Biol*. **26**(13), 5096-5105 (2006). PMID: 1489176.
11. C. S. Osborne, L. Chakalova, J. A. Mitchell, A. Horton, A. L. Wood, D. J. Bolland, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*. **5**(8), e192 (2007). PMID: 1945077.
12. F. Bantignies, V. Roure, I. Comet, B. Leblanc, B. Schuettengruber, J. Bonnet, et al. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell*. **144**(2), 214-226 (2011).
13. C. H. Eskiw, N. F. Cope, I. Clay, S. Schoenfelder, T. Nagano and P. Fraser. Transcription factories and nuclear organization of the genome. *Cold Spring Harb Symp Quant Biol*. **75**, 501-506 (2010).
14. T. Misteli. Higher-order genome organization in human disease. *Cold Spring Harb Perspect Biol*. **2**(8), a000794 (2010).
15. J. J. Roix, P. G. McQueen, P. J. Munson, L. A. Parada and T. Misteli. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet*. **34**(3), 287-291 (2003).
16. C. Lin, L. Yang, B. Tanasa, K. Hutt, B. G. Ju, K. Ohgi, et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*. **139**(6), 1069-1083 (2009). PMID: 2812435.
17. R. S. Mani, S. A. Tomlins, K. Callahan, A. Ghosh, M. K. Nyati, S. Varambally, et al. Induced chromosomal proximity and gene fusions in prostate cancer. *Science*. **326**(5957), 1230 (2009). PMID: 2935583.
18. F. Hediger, F. R. Neumann, G. Van Houwe, K. Dubrana and S. M. Gasser. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr Biol*. **12**(24), 2076-2089 (2002).
19. A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*. **3**(5), e157 (2005). PMID: 1084335.
20. D. P. Bazett-Jones, R. Li, E. Fussner, R. Nisman, and H. Dehghani. Elucidating Chromatin and Nuclear Domain Architecture with Electron Spectroscopic Imaging. *Chromosome Res*. **16**, 397-412 (2008).
21. L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. **453**(7197), 948-951 (2008).
22. M. J. Fullwood, M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. B. Mohamed, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. **462**(7269), 58-64 (2009). PMID: 2774924.

23. L. Handoko, H. Xu, G. Li, C. Y. Ngan, E. Chew, M. Schnapp, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet.* **43**(7), 630-638 (2011).
24. J. Dekker, K. Rippe, M. Dekker, N. Kleckner. Capturing chromosome conformation. *Science.* **295**(5558), 1306-1311 (2002).
25. J. Dostie, T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**(10), 1299-1309 (2006). PMID: 1581439.
26. S. Horike, S. Cai, M. Miyano, J. F. Cheng and T. Kohwi-Shigematsu. Loss of silent chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet.* **37**(1), 31-40 (2005).
27. J. Q. Ling, T. Li, J. F. Hu, T. H. Vu, H. L. Chen, X. W. Qiu, et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science.* **312**(5771), 269-272 (2006).
28. S. Lomvardas, G. Barnea, D. J. Pisapia, M. Mendelsohn, J. Kirkland and R. Axel. Interchromosomal interactions and olfactory receptor choice. *Cell.* **126**(2), 403-413 (2006).
29. M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet.* **38**(11), 1348-1354 (2006).
30. V. K. Tiwari, L. Cope, K. M. McGarvey, J. E. Ohm and S. B. Baylin. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.* **18**(7), 1171-1179 (2008). PMID: 2493406.
31. Z. Zhao, G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano, S. Wang, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* **38**(11), 1341-1347 (2006).
32. D. Bau, A. Sanyal, B. R. Lajoie, E. Capriotti, M. Byron, J. B. Lawrence, et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol.* **18**(1), 107-114 (2011). PMID: 3056208.
33. S. Schoenfelder, T. Sexton, L. Chakalova, N. F. Cope, A. Horton, S. Andrews, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet.* **42**(1), 53-61 (2010).
34. B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell.* **10**(6), 1453-1465 (2002).
35. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* **326**(5950), 289-293 (2009).
36. Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau, and W. S. Noble. A three-dimensional model of the yeast genome. *Nature.* **465**, 363-367 (2010). PMID: 2874121.
37. H. Tanizawa, O. Iwasaki, A. Tanaka, J. R. Capizzi, P. Wickramasignhe, M. Lee, Z. Fu, and K. Noma. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucl Acids Res.* (2010).

38. C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol.* **7**, 485(2011).
39. L. S. Churchman and J. S. Weissman. Nascent transcript sequencing visualizes transcription at nucleotide level. *Nature.* **469**(7330), 368-373 (2011).
40. A. J. Jasiak, H. Hartmann, E. Karakasili, M. Kalocsay, A. Flatley, E. Kremmer, K. Strässer, D. E. Martin, J. Söding and P. Cramer. Genome-associated RNA Polymerase II Includes the Dissociable Rpb4/7 Subcomplex. *J Biol Chem.* **283**(39), 26423-26427 (2008).
41. D. Bau, A. Sanyal, B. R. Lajoie, E. Capriotti, M. Byron, J. B. Lawrence, J. Dekker, and M. A. Marti-Renom. The three-dimensional folding of the  $\alpha$ -globin gene domain reveals formation of chromatin globules. *Nature Structural Molecular Biology.* **18**(1), 107–114 (2011).
42. M. A. Marti-Renom and L. Mirny. Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS Computational Biology.* **7**(7), e100002125 (2011).
43. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler. The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
44. P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, et al. Ensembl 2011. *Nucleic Acids Research.* **39**, D800-D806 (2010).
45. M. E. Skinner, A. V. Uzilov, L. D. Stein, C. J. Mungall and I. H. Holmes. JBrowse: a next-generation genome browser. *Genome Res.* **19**(9), 1630-1638 (2009).
46. T. M. Asbury, M. Mitman, J. Tang, and W. J. Zheng. Genome3D: a viewer-model framework for integrating and visualizing multi-scale epigenomic information within a three-dimensional genome. *BMC Bioinformatics.* **11**, 444 (2010).
47. Z. Dai and X. Dai. Nuclear colocalization of transcription factor target genes strengthens coregulation in yeast. *Nucl Acids Res.* (2011).