**doi** 10.18699/bgrs2026-130

# От геномики к хроматомике: насколько мы близки к пониманию того, как функционируют эукариотические геномы?

Шайтан А.К.[1, 2*]
[1] *МГУ имени М.В. Ломоносова, Москва, Россия*
[2] *НИУ Высшая школа экономики, Москва, Россия*
** shaytanak@my.msu.ru*

**Ключевые слова:** хроматин; биофизика; геномика; моделирование



Проблема понимания функционирования человеческого и других эукариотических геномов является одной из ключевых задач молекулярной биологии, биоинформатики и биофизики последних десятилетий. Огромную роль в исследовании геномов сыграли и продолжают играть методы геномики, основанные на технологиях секвенирования. Успех этих методов во многом обусловлен тем фундаментальным фактом, что живые системы используют дискретную цифровую основу для кодирования и передачи генетической информации в виде последовательностей ДНК/РНК и белков. Однако большая часть процессов регуляции работы геномов и связанная с ними обработка информационных сигналов обусловлены уже «аналоговыми» физическими взаимодействиями молекул в физическом носителе генома, коим является хроматин, динамический комплекс молекул ДНК, РНК и белков в ядрах клеток.

Медленный прогресс в целостном понимании работы хроматина обусловлен сложностью исследования и построения динамических моделей трехмерной структуры генома. С экспериментальной точки зрения ядра клеток находятся в «неудобном» масштабном диапазоне, где методы микроскопии уже не дают приемлемой детализации структуры объектов, а методы структурной биологии еще не способны исследовать структуры такого масштаба. С точки зрения построения моделей процессов работы хроматина сложность заключается в многокомпонентности системы, сложности параметризации моделей взаимодействий и их нелинейности. Ситуация осложняется динамичностью и мультивалентностью взаимодействий в хроматине, активным АТФ-зависимым ремоделированием структуры и топологии хроматина, макромолекулярным краудингом с активным вовлечением неупорядоченных регионов белков. По нашим оценкам, в среднем около половины длины последовательностей белков хроматина приходится на неупорядоченные регионы [1]. Наш недавний метаанализ различных экспериментальных данных о белках хроматина человека указывает на то, что вопрос о количестве типов белков и их представленности в ядрах клеток остается далеко не полностью разрешенным [1].

Тем не менее можно констатировать, что наше понимание функционирования хроматина за последние годы значительно улучшилось, в том числе благодаря развитию ряда биофизических, молекулярно-биологических и вычислительных методов моделирования. Существенный прогресс в области криоэлектронной микроскопии позволяет «прощупывать» супрануклеосомную структуру хроматина [2, 3], а методы мультимасштабного моделирования позволяют оценивать, каким образом функционирование генома осуществляется в рамках его компактной упаковки в хроматиновые фибриллы [4]. Применение методов машинного обучения выводит на новый уровень возможности по соотнесению последовательностей ДНК/РНК/белков с их функциональными эффектами и трехмерной структурой. В частности, для структурированных глобулярных доменов белков хроматина удается предсказывать ранее неизвестные белковые домены [1] и структуры взаимодействующих белковых комплексов, а для неструктурированных фрагментов соотносить их аминокислотный состав со сродством белка к различным немембранным органеллам.

В данном докладе я постараюсь дать актуальную картину функционирования эукариотических геномов, которая формируется на стыке геномных и эпигеномных исследований, биофизических экспериментов и моделирования, а также подчеркнуть актуальные направления дальнейших исследований.

*Финансирование*: Исследование поддержано РНФ (грант № 25-14-00046).

# From genomics to chromatomics: how close are we to understanding how eukaryotic genomes function?

Shaytan A.K.[1, 2*]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *HSE University, Moscow, Russia*
** shaytanak@my.msu.ru*

**Key words:** chromatin; biophysics; genomics; modelling

The problem of understanding the functioning of the human genome and other eukaryotic genomes has been one of the central challenges of molecular biology, bioinformatics, and biophysics in recent decades. Genomics methods based on sequencing technologies have played, and continue to play, a tremendous role in genome research. The success of these methods is largely due to the fundamental fact that living systems use a discrete digital basis for encoding and

transmitting genetic information in the form of DNA/RNA and protein sequences. However, a large proportion of the processes involved in genome regulation, as well as the associated processing of informational signals, are determined by the "analog" physical interactions of molecules within the physical carrier of the genome, namely chromatin, a dynamic complex of DNA, RNA, and proteins in cell nuclei.

Slow progress toward a comprehensive understanding of chromatin function is due to the difficulty of both experimental investigation and the construction of dynamic models of the three-dimensional genome structure. From an experimental point of view, cell nuclei fall within an "inconvenient" size range in which microscopy methods no longer provide sufficient structural detail, while structural biology methods are still unable to investigate objects on such a scale. From the perspective of modeling chromatin function, the main difficulties lie in the multicomponent nature of the system, the complexity of parameterizing interaction models, and their nonlinearity. The situation is further complicated by the dynamic and multivalent nature of interactions in chromatin, active ATP-dependent remodeling of chromatin structure and topology, and macromolecular crowding involving the active participation of intrinsically disordered protein regions. According to our estimates, on average about half of the sequence length of chromatin proteins consists of disordered regions [1]. Our recent meta-analysis of diverse experimental data on human chromatin proteins indicates that the question of how many types of proteins are present in cell nuclei, and in what abundance, is still far from fully resolved [1].

Nevertheless, it can be stated that our understanding of chromatin function has improved substantially in recent years, including through the development of a range of biophysical, molecular biological, and computational modeling methods. Significant progress in cryo-electron microscopy has made it possible to probe the supranucleosomal structure of chromatin [2, 3], while multiscale modeling methods make it possible to assess how genome function is carried out within the framework of its compact packaging into chromatin fibrils [4]. The application of machine learning methods has brought to a new level our ability to relate DNA/RNA/protein sequences to their functional effects and three-dimensional structure. In particular, for structured globular domains of chromatin proteins, it has become possible to predict previously unknown protein domains [1] and the structures of interacting protein complexes, while for unstructured fragments it has become possible to relate their amino acid composition to the affinity of a protein for various membraneless organelles.

In this lecture, I will attempt to provide an up-to-date picture of the functioning of eukaryotic genomes as it is emerging at the intersection of genomic and epigenomic research, biophysical experiments, and modeling, and to highlight promising directions for future research.

*Funding*: The study is supported by Russian Science Foundation (Grant No. 25-14-00046).

*Список литературы/References*

1. Gribkova, A.K.; Armeev, G.A.; Kirpichnikov, M.P.; Shaytan, A.K. (Re)Defining the Human Chromatome: An Integrated Meta-Analysis of Localization, Function, Abundance, Physical Properties, and Domain Composition of Chromatin Proteins. Nucleic Acids Research 2026, 54, gkaf1489, doi:10.1093/nar/gkaf1489.
2. Armeev, G.A.; Gribkova, A.K.; Shaytan, A.K. Nucleosomes and Their Complexes in the cryoEM Era: Trends and Limitations. *Front. Mol. Biosci.* **2022**, *9*, 1070489, doi:10.3389/fmolb.2022.1070489.
3. Armeev, G.A.; Gribkova, A.K.; Shaytan, A.K. NucleosomeDB - a Database of 3D Nucleosome Structures and Their Complexes with Comparative Analysis Toolkit 2023.
4. Fedulova, A.S.; Armeev, G.A.; Romanova, T.A.; Singh-Palchevskaia, L.; Kosarim, N.A.; Motorin, N.A.; Komarova, G.A.; Shaytan, A.K. Molecular Dynamics Simulations of Nucleosomes Are Coming of Age. WIREs Computational Molecular Science 2024, 14, e1728, doi:10.1002/wcms.1728.