

Алгоритм кластеризации и сравнения иерархий в пространственной структуре хроматина

Научный руководитель – Рубинов Анатолий Рувимович

Пушкарёва Ольга Андреевна

Студент (бакалавр)

Национальный исследовательский университет «Высшая школа экономики», Факультет математики, Москва, Россия

E-mail: mrs.push@gmail.com

В последние годы широкое распространение в изучении хроматина получили методы фиксации конформации хромосом (3C, 4C, 5C, Hi-C), известные под общим названием C-методов. Эти методы позволили взглянуть на трехмерную структуру хроматина и показать, что она играет важную роль в регуляции его активности. Одним из наиболее важных достижений стала демонстрация того, что геном разбивается на так называемые Топологически Ассоциированные Домены (ТАДы). ТАДы характеризуются высокой частотой контактов между удаленными участками генома внутри себя, и пониженной, если элементы находятся в разных доменах. Было придумано множество методов, позволяющих искать ТАДы (Armatus, Insulation score, TadBit, Arrowhead etc). Однако, все эти методы зависят от разных произвольных параметров, что приводит к непохожим результатам выделения доменов у разных алгоритмов. Основываясь на алгоритме, предложенном А.Р. Рубиновым (2017), мы предлагаем не зависящий от параметров метод анализа Hi-C данных, позволяющий построить дерево иерархий на бинированной карте и сравнить полученные деревья для выбранных геномов.

Источники и литература

- 1) Разин С. В., Гаврилов А.А., Кос П., Ульянов С.В. Самоорганизация хроматиновой фибриллы в топологически-ассоциированные домены // Биоорганическая химия, 2017, том 43, № 2, с. 115–123
- 2) Рубинов А. Р. “1-D hierarchical clustering”, 2017 (неопубликованное)
- 3) Forcato M., Nicoletti C., Pal K., Livi C.M., Ferrari F., Bicciato S. Comparison of computational methods for Hi-C data analysis, Nature Methods volume 14, 2017, pages 679–685