

Статистический анализ компартментов в данных Hi-C.

Научный руководитель – Миронов Андрей Александрович

Сефербекова Заира Назимовна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия

E-mail: sef.zaira@gmail.com

Пространственная организация генома эукариот играет важную роль в регуляции транскрипции генов [1, 2]. Появление высокопроизводительного метода Hi-C позволило продвинуться в изучении организации генома в различных типах клеток [4, 5].

Анализ данных, полученных в экспериментах Hi-C, показал, что весь геном может быть поделен на два компартмента А и В таких, что контакты внутри каждого из них встречаются чаще, чем контакты между ними [4]. Считается, что в компартмент А входит транскрипционно активный хроматин, а в компартмент В - транскрипционно неактивный хроматин [4].

Несмотря на то, что другие методы находят большее число компартментов [5, 6] или не находят их вовсе [3], описанный подход [4] не потерял актуальности. Тем не менее, он обладает рядом недостатков: 1) при делении на компартменты учитывается знак главной компоненты, но не ее значение, которое варьирует между локусами, 2) наибольший вклад в компартменты могут вносить близкие контакты, 3) не оценивается статистическая значимость компартмента. Целью настоящей работы было создание модели для построения фонового распределения для оценки статистической значимости принадлежности локуса компартменту и анализ вклада близких контактов в компартменты.

Мы показали, что разделение на компартменты для некоторых хромосом сохраняется даже после удаления значительной части матрицы контактов. При этом тенденция к сохранению контактов была более выражена у линии GM06990, чем у раковой линии K562. С одной стороны, это может говорить о том, что компартменты основаны преимущественно на дальних взаимодействиях. С другой стороны, сохранение компартментов после удаления большей части матрицы может быть обусловлено тем, что используемый математический метод всегда может выделить какие-то компартменты.

В работе также предложены две модели для оценки статистической значимости компартментов. Первая модель использует в качестве фонового распределения матрицы, полученные из исходной путем перемешивания элементов внутри диагоналей, а вторая - матрицы с частью элементов исходной матрицы и остальными элементами, равными а) нулю или б) среднему значению компоненты по диагонали. Разделение на компартменты можно производить с использованием предложенных нами моделей, что позволит оценивать статистическую значимость компонент, наблюдаемых в реальных Hi-C данных.

Источники и литература

- 1) Bickmore W.A. The spatial organization of the human genome // Annu. Rev. Genomics Hum. Genet. 14, 2013, pp. 67–84.
- 2) Cremer T. et al. Chromosome territories, nuclear architecture and gene regulation in mammalian cells // Nat. Rev. Genet. 2, 2001, pp. 292–301.

- 3) Imakaev M. et al. Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization // Nat. Methods 9(10), 2012, pp. 999–1003.
- 4) Lieberman-Aiden E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome // Science 326, 2009, pp. 289–293.
- 5) Rao S. S. P. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping // Cell 159, 2014, pp.1665–1680.
- 6) Yaffe E. et al. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture // Nat. Genet. 43(11), 2011, doi:10.1038/ng.947.