

# ИДЕНТИФИКАЦИЯ СТРУКТУРНЫХ ВАРИАЦИЙ В ГЕНОМЕ КЛЕТОК ОСТРОГО МИЕЛОИДНОГО ЛЕЙКОЗА ПО ДАННЫМ CAPTURE Hi-C

М.А. Лиховец, Е.В. Гузова, Н.Н. Яцков, В.В. Скакун, В.В. Гринев

*Белорусский государственный университет, Минск, Беларусь  
E-mail: likhovetsmax@gmail.com*

Структурные вариации генома могут быть важными драйверными мутациями онкогенеза. В то же время их обнаружение по-прежнему является сложной задачей. В настоящей работе представлены результаты поиска структурных вариаций в геноме клеток модельной линии острого миелоидного лейкоза человека Kasumi-1 по данным Capture Hi-C. На основании этих данных были получены координаты потенциальных перестроек в геноме изучаемых клеток. Для части обнаруженных вариаций получены подтверждения их транскрипционной активности.

**Ключевые слова:** Capture Hi-C; 3D структура хроматина; структурные вариации; транскрипция; лейкозные клетки человека.

## ВВЕДЕНИЕ

В последние десятилетия наблюдается активное развитие методов определения 3D структуры хроматина с использованием секвенирования нового поколения. Наиболее популярными и одними из самых мощных среди них являются методы, основанные на технологии высокопроизводительного захвата конформации хромосом, или Hi-C. Данная технология позволяет получить частоты всех попарных контактов между участками хроматина в ядре. Как правило, изначально такие данные получают для изучения особенностей архитектуры генома и обнаружения различных регуляторных взаимодействий. Однако было показано, что данные Hi-C можно использовать для таких задач, как сборка генома или обнаружение протяженных мутаций – структурных вариаций. Структурные вариации включают делеции, инверсии, дупликации и транслокации. Они могут приводить к изменению копийности генов, менять регуляцию их экспрессии, а также приводить к образованию гибридных (онко)генов [1].

## ОБНАРУЖЕНИЕ СТРУКТУРНЫХ ВАРИАЦИЙ

В настоящей работе использовались данные Capture Hi-C для клеток модельной линии острого миелоидного лейкоза человека Kasumi-1. Этот метод представляет собой модификацию Hi-C, при которой библиотека для секвенирования обогащается фрагментами, содержащими промоторы интересующих генов (всего более 20000 различных промоторов) [2].

В дальнейшем при обработке данных использовались координаты захваченных при процедуре Capture Hi-C фрагментов, которые были разработаны

по эталонному геному hg19. В связи с этим изначально данные Capture Hi-C картировались на геном hg19, а обработанные результаты были переведены в координаты hg38 при помощи HiCLift [3]. Парноконцевые прочтения были получены из SRA с кодами доступа SRR7521620 и SRR7521621. Предварительная обработка, картирование на эталонный геном, фильтрация артефактов и дубликатов полимеразной цепной реакции была проведена при помощи пайплайна HiCUP [4]. Из полученных пар взаимодействующих фрагментов с помощью Python пакета cooler рассчитывалась матрица контактов с разрешением 10Kb [5]. Учитывая особенности процедуры Capture Hi-C, была также проведена нормализация матрицы контактов, что необходимо для надежного предсказания структурных вариаций. Кроме того, картированные и обработанные прочтения были использованы для определения существенных взаимодействий между участками генома, успешно проведенного благодаря функционалу R/Bioconductor пакета CHiCAGO [6].

Матрицы контактов Hi-C могут быть визуализированы в виде тепловых карт, где частоты взаимодействия между двумя участками хроматина отображаются интенсивностью окраски соответствующего пикселя. Для идентификации различных структурных вариаций была использована сверточная нейронная сеть EagleC, способная классифицировать паттерны на картах Hi-C, соответствующие вариациям структуры [7]. Ключевым результатом работы EagleC является список геномных координат мест разрыва-соединения хромосом(-ы), или так называемых брейкпоинтов (BP, от англ. breakpoint) и их ориентация.

В клетках Kasumi-1 с помощью EagleC было идентифицировано множество BPs, которые в дальнейшем были отфильтрованы против событий, попадающих в области центромер, а также событий, имеющих низкую подтверждаемость. Всего после фильтрации осталось 16 BPs, координаты части из которых представлены в таблице. Эти координаты соответствуют эталонному геному hg38 и имеют разрешение +/- 10Kb.

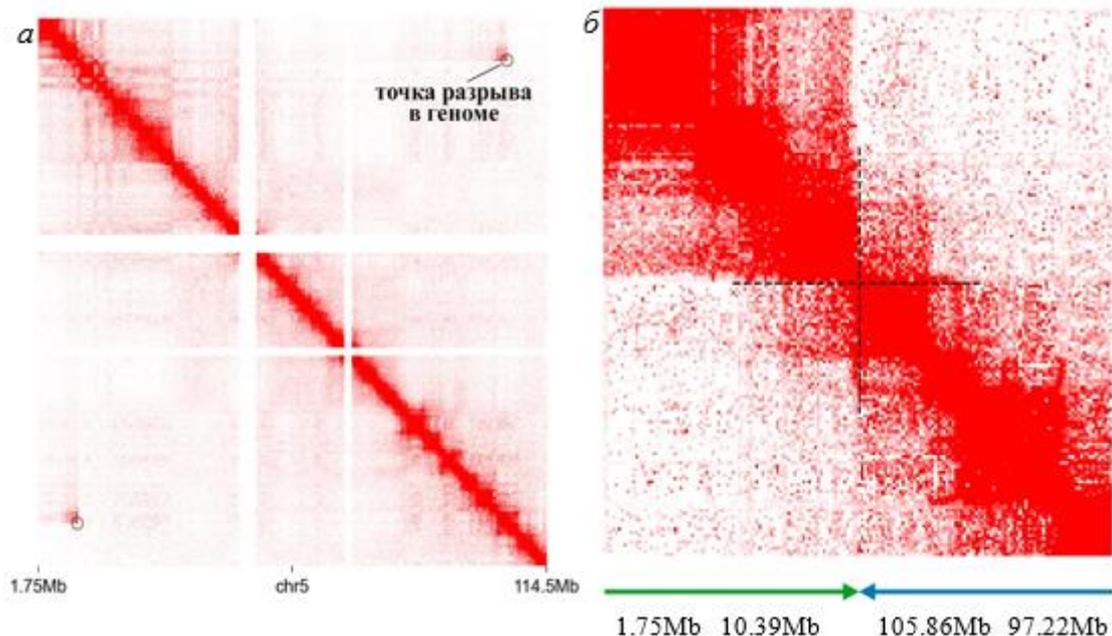
Полученные координаты точек разрыва соотносятся с опубликованными ранее данными о структурных вариациях в клетках Kasumi-1 [8]. Кроме того, достоверность полученных предсказаний подтверждается обнаружением известной для этих клеток реципрокной транслокации t(8;21)(q22;q22). Однако ряд BPs, обнаруженных в представленной работе, не был ранее описан.

В завершение нами была предпринята попытка выяснить ведут ли обнаруженные нами геномные перестройки к появлению новых транскрипционно активных гибридных (онко)генов.

### Координаты наиболее значимых ВРs в геноме клеток Kasumi-1

Хромо-сома ВР № 1	Координата ВР № 1	Хромо-сома ВР № 2	Координата ВР № 2	Вероятность ориентации			
				3'-3'	3'-5'	5'-3'	5'-5'
15	46377802	9	29300002	4,6e-19	3,7e-24	9,4e-28	1,0
2	203465277	8	122677761	1,1e-15	1,0	7,5e-11	2,7e-13
21	34857703	8	92047772	1,9e-20	1,0	0,9	7,0e-19
16	16776143	16	18086143	4,0e-20	1,7e-16	1,0	6,5e-19
5	10389888	5	105894299	1,0	1,1e-24	1,5e-15	2,1e-14
3	113211153	3	114471153	3,9e-15	3,5e-10	1,0	4,2e-14

Для этого мы сопоставили геномные координаты ВРs, идентифицированных по данным Capture Hi-C, с координатами ВРs, которые были обнаружены нами ранее по данным полнотранскриптомного секвенирования RNA-Seq. Учитывая разрешение матриц контактов и потенциальное отличие координат ВРs из RNA-seq из-за сплайсинга, мы провели поиск ВРs из RNA-Seq, координаты которых находятся в пределах +/- 50Kb от соответствующего ВР из данных Capture Hi-C. В результате был найден транскрипт гибридного онкогена RUNX1-RUNX1T1, соответствующий известной реципрокной транслокации t(8;21)(q22;q22), а также транскрипт, соответствующий ВР на хромосоме 5. Данный ВР ранее не был описан, его координаты соответствуют гену MARCHF6 и межгенной области (рисунок).



Карты контактов Hi-C для участка хромосомы 5 без учета структурной вариации (а) и с учетом (б)

## ЗАКЛЮЧЕНИЕ

Таким образом, на основании данных о пространственной организации хроматина были получены координаты потенциальных ВРs в геноме клеток Kasumi-1. Одна из вновь обнаруженных точек рекомбинации, согласно данным транскриптомного секвенирования, приводит к образованию гибридного (онко)гена, состоящего из двух удалённых участков генома. Тем не менее, представленные в данной работе результаты требуют последующей валидации экспериментальными методами.

Особый интерес представляет более подробное изучение изменения структуры хроматина в областях соответствующих перестроек и изучение влияния такой структуры на экспрессию расположенных рядом генов. Кроме того, полученные данные являются основой для дальнейшего поиска ассоциаций между гибридными транскриптами, обнаруживаемыми по данным RNA-Seq, и 3D организацией хроматина.

## БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. / Wang X., Xu J., Zhang B. [et al.] // *Nat. Methods*. 2021. Vol. 18, № 6. P. 661-668. DOI: 10.1038/s41592-021-01164-w
2. RUNX1-ETO depletion in t(8;21) AML leads to C/EBP $\alpha$ - and AP-1-mediated alterations in enhancer-promoter interaction. / Ptasinska A., Pickin A., Assi S.A. [et al.] // *Cell Rep*. 2019. Vol. 28, № 12. P. 3022-3031.e7. DOI: 10.1016/j.celrep.2019.08.040
3. Wang X., Yue F. HiCLift: a fast and efficient tool for converting chromatin interaction data between genome assemblies. // *Bioinformatics*. 2023. Vol. 39, № 6. P. btad389. DOI: 10.1093/bioinformatics/btad389
4. Wingett S.W., Ewels P., Furlan-Magaril M. HiCUP: pipeline for mapping and processing Hi-C data. // *F1000Research*. 2015. Vol. 4. P. 1310. DOI: 10.12688/f1000research.7334.1
5. Abdennur N., Mirny L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. // *Bioinformatics*. 2020. Vol. 36, № 1. P. 311-316. DOI: 10.1093/bioinformatics/btz540
6. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. / Cairns J., Freire-Pritchett P., Wingett S.W. [et al.] // *Genome Biol*. 2016. Vol. 17, № 1. P. 127. DOI: 10.1186/s13059-016-0992-2
7. Wang X., Luan Y., Yue F. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. // *Sci. Adv*. 2022. Vol. 8, № 24. P. eabn9215. DOI: 10.1126/sciadv.abn9215
8. Subtype-specific 3D genome alteration in acute myeloid leukaemia. / Xu J., Song F., Lyu H. [et al.] // *Nature*. 2022. Vol. 611, № 7935. P. 387-398. DOI: 10.1038/s41586-022-05365-x