

Нанопоровое секвенирование конкатемерных библиотек вариабельных областей TCR из опухолевой ткани колоректального рака

Котова Е.С.^{1*}, Петрова Т.В.¹, Бабенко В.В.¹, Болдырева Д.И.¹, Гарманова Т.Н.²,
Лазарев В.Н.¹, Шарова Е.И.¹

¹ ФНКЦ ФХМ имени академика Ю.М. Лопухина ФМБА России, Москва, Россия

² Московский государственный университет им. М.В. Ломоносова, Москва, Россия

* potamanthus@gmail.com

Ключевые слова: нанопоровое секвенирование; конкатемеры; RCA; TCR; колоректальный рак

Мотивация и цель: Секвенирование нового поколения (NGS-Next Generation Sequencing) позволило провести множество важных исследований в области молекулярной онкологии. Важно отметить, что методы NGS все еще находятся в стадии разработки и имеют свои ограничения. Секвенаторы с высоким качеством прочтений не могут производить прочтения достаточной длины, а секвенаторы с длинными прочтениями не могут обеспечить прочтения высокой точности с приемлемым выходом. Нанопоровое секвенирование конкатемерных библиотек с получением консенсуса мономерных последовательностей позволяет получать достаточно длинные считывания с более высокой точностью [1, 2]. Мы предлагаем новый метод получения конкатемерных библиотек ампликонов без использования праймеров на этапе получения конкатемеров путем амплификации по типу катящегося кольца (RCA-Rolling Circle Amplification), для уменьшения доли химерных молекул. Мы провели предварительное тестирование разработанного протокола.

Методы и алгоритмы: С помощью набора Native Barcoding Kit 24 V14 были получены шесть баркодированных библиотек, соответствующих вариабельным областям альфа-цепи Т-клеточного рецептора (TCRA – T-cell receptor alpha chain) и бета-цепи Т-клеточного рецептора (TCRB – T-cell receptor alpha chain) из трех образцов РНК опухолей колоректального рака. Каждый мономер в конкатемерной молекуле нес уникальный молекулярный индекс (УМИ) длиной 17 п.о. для обнаружения и фильтрации химерных прочтений. Секвенирование проводили на проточной ячейке R10.4. Полученные данные анализировали с помощью пакетов Bowtie2, MAFFT, MiXCR и собственных скриптов, написанных на bash, питоне 3 и R.

Результаты: Для библиотек вариабельных участков TCRA было получено 3954-6668 прочтений, TCRB – 4013-6427. Медианное количество мономеров для всех библиотек составило 7; 25 % прочтений содержали более 13–14 мономерных единиц, а 10 % – более 22–24. Консенсусные последовательности были получены для каждого конкатемера и проанализированы с помощью программы MiXCR. Мы идентифицировали 33–69 и 29–176 клонов TCRA и TCRB соответственно. Клонов, которым соответствовали, как минимум, 3 прочтения, было 3–21 (TCRA) и 4–14 (TCRB). Количество считываний для наиболее представленных клонов составило 31 (TCRA) и 22 (TCRB).

Выводы: Применение метода RCA без использования праймеров с последующим нанопоровым секвенированием позволило успешно реконструировать вариабельные области TCRA и TCRB полной длины на основе нуклеотидных последовательностей мономеров, несущих УМИ. Первичные данные свидетельствуют о том, что это перспективный подход к полноразмерному секвенированию и высокопроизводительному профилированию гетерогенных РНК с высоким уровнем сходства, к примеру, РНК TCR, BCR и ретроэлементов в опухолевых образцах.

Финансирование: Данная публикация выполнена в рамках государственного задания «Персональный онкопрепарат», номер государственного учета НИОКТР 124031200002-3.

Nanopore concatemeric sequencing of TCR variable regions from colorectal tumors

Kotova E.S.^{1*}, Petrova T.V.¹, Babenko V.V.¹, Boldyreva D.I.¹, Garmanova T.N.², Lazarev V.N.¹, Sharova E.I.¹

¹ Lopukhin FRCC PCM, Moscow, Russia

² Moscow State University, Moscow, Russia, Moscow, Russia

* potamanthus@gmail.com

Key words: nanopore sequencing; concatemers; RCA; TCR; colorectal cancer

Motivation and Aim: Next-generation sequencing has enabled many important studies in molecular oncology. It is important to note that this technique is still in development and has limitations. Sequencers with high read quality cannot provide reads that are long enough, and long-read sequencers cannot provide reads that are accurate enough with acceptable yields. Nanopore sequencing of concatemeric libraries, with the derivation of monomeric sequence consensus, allows for the obtainment of sufficiently long reads with higher accuracy [1, 2]. We propose a new method for obtaining concatemeric libraries of amplicons without primers during the concatemer production stage by using rolling circle amplification (RCA) to decrease the portion of chimeric molecules. We conducted a preliminary test of the developed protocol.

Methods and Algorithms: Six barcoded libraries were created using the Native Barcoding Kit 24 V14, corresponding to the variable regions of TCRA (T-cell receptor alpha chain) and TCRB (T-cell receptor beta chain) from three RNA samples of colorectal cancer tumors. Each monomer in the concatemeric molecule was labeled with a 17 bp unique molecular identifier (UMI) to detect and filter out chimeric reads. Sequencing was performed using the R10.4 flow cell. We analyzed obtained data using packages Bowtie2, MAFFT, MiXCR and our own bash-, python3- and R-scripts.

Results: The TCRA libraries yielded 3954-6668 reads, while the TCRB libraries yielded 4013-6427 reads. The median number of monomeric units for all libraries was 7; 25 % of the reads contained more than 13-14 monomeric units, and 10 % contained more than 22-24. Consensus sequences were obtained for each concatemer and analyzed using the Mixcr software. We identified 33-69 and 29-176 TCRA and TCRB clones respectively. The clones that matched at least 3 reads were 3-21 (TCRA) and 4-14 (TCRB). The number of reads for the most represented clones was 31 (TCRA) and 22 (TCRB).

Conclusion: The entire length of the TCRA and TCRB variable regions were successfully reconstructed using UMI-labeled monomers through a primerless RCA approach and nanopore sequencing. The initial data suggests that this approach is very promising for whole length sequencing and high throughput profiling of heterogeneous RNAs with high similarity, such as TCR, BCR, and retroelements in tumor samples.

Funding: This publication was carried out with state funding for the «Personalnyi onkopreparat» project, state registration number of R&D 124031200002-3.

Список литературы/References

- 1 Li C. et al. INC-Seq: accurate single molecule reads using nanopore sequencing. *Gigascience*. 2016;5(1):34
- 2 Volden R. et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci USA*. 2018;115(39):9726-9731