

Деконволюция клеточного состава тканей на основе данных секвенирования микроРНК

Зарубин А.А., Марков А.В., Слепцов А.А., Назаренко М.С.

Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук
634050, г. Томск, ул. Набережная Ушайки, 10

Экспрессия микроРНК (miRNA) изменяется под действием различных факторов, что создаёт проблему интерпретации результатов исследований, поскольку патологические процессы могут изменять не только функциональную активность клеток, но и соотношения клеточных популяций. Возможным решением является применение подходов клеточной деконволюции. Нами был разработан алгоритм создания референсной панели экспрессии miRNA, специфичной для определенных типов клеток, что позволяет оценить представленность различных клеточных популяций в стенке артерий. Показано, что существующие алгоритмы деконволюции подходят для данных miRNA.

Ключевые слова: микроРНК, деконволюция, секвенирование.

Для цитирования: Зарубин А.А., Марков А.В., Слепцов А.А., Назаренко М.С. Деконволюция клеточного состава тканей на основе данных секвенирования микроРНК. *Медицинская генетика* 2020; 19(12): 64-65.

DOI: 10.25557/2073-7998.2020.12.64-65

Автор для корреспонденции: Зарубин Алексей Андреевич; **e-mail:** aleksei.zarubin@medgenetics.ru

Финансирование. Программа фундаментальных научных исследований РАН.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила: 20.05.2020.

Tissue-specific deconvolution of cell composition using miRNA sequencing data

Zarubin A.A., Markov A.V., Sleptcov A.A., Nazarenko M.S.

Research Institute of Medical Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences
Naberezhnaya Ushaiki 10, Tomsk, 634050, Russia

The expression of miRNAs is affected by different factors. This poses challenges when interpreting the results of the study. Pathological processes can influence both the functional activity of cells and the ratio of cell populations. A possible solution is to utilize the approaches of deconvolution. We have developed an algorithm to assess reference panel from the microRNA dataset on the example of artery walls. In this study, we have shown that existing deconvolution algorithms are suitable for miRNA data.

Keywords: miRNA, deconvolution, sequencing.

For citation: Zarubin A.A., Markov A.V., Sleptcov A.A., Nazarenko M.S. Tissue-specific deconvolution of cell composition using miRNA sequencing data. *Medical genetics* 2020; 19(12): 64-65. (In Rus.).

DOI: 10.25557/2073-7998.2020.12.64-65

Corresponding author: Zarubin Alexei; **e-mail:** aleksei.zarubin@medgenetics.ru

Funding. Siberian branch of the Russian Academy of Sciences, the Program for basic scientific research.

Conflict of interest. Authors declare no conflict of interest

Accepted: 20.05.2020

МикроРНК (miRNA) — это малые РНК, регулирующие экспрессию генов, в основном, на посттранскрипционном уровне. Изменение уровня экспрессии miRNA показано при различных патологических процессах. Помимо этого, экспрессия miRNA тканеспецифична. Эта особенность создаёт проблему интерпретации результатов исследования дифференциальной экспрессии miRNA, поскольку патологический процесс может изменять клеточный состав. При определении экспрессии мРНК данная про-

блема решается посредством деконволюции сигналов экспрессии генов по типам клеток, что позволяет оценить клеточный состав исследуемого образца. Клеточная деконволюция — это набор статистических методов, которые позволяют восстановить информацию о соотношении различных клеточных линий в образце. Алгоритмы деконволюции классифицируются по типу используемых данных и по необходимости использования референсного набора данных с информацией, полученной из образцов «чистых» клеточных ли-

ний. Однако методы деконволюции имеют технические ограничения: максимальная точность и разрешающая способность метода зависит от количества определяемых сигналов. Это ограничение уменьшает возможность использования методов деконволюции к данным miRNA секвенирования, поскольку количество получаемых маркеров при miRNA секвенировании значительно меньше, чем при оценке метилирования ДНК или экспрессии генов. Вместе с тем, многочисленные базы данных по экспрессии miRNA и новые алгоритмы потенциально позволяют разрешить и эту задачу.

Цель: оценить эффективность использования существующих биоинформатических подходов клеточной деконволюции в отношении данных секвенирования miRNA и разработать алгоритм создания референсной панели экспрессии miRNA, специфичной для клеток, на примере стенки артерии с лейкоцитарной инфильтрацией.

Материалы и методы

При создании референсных паттернов экспрессии miRNA клеточных линий использовались данные атласа экспрессии miRNA проекта FANTOM5 [1]: эндотелиальные клетки артерий (ЭК; n=9), фибробласты аорты (ФБ; n=3), гладкомышечные клетки артерий (ГМК; n=9), макрофаги (n=3), нейтрофилы (n=3), В- (n=4), Т-(CD8- и CD4- (n=6)) и NK-клетки (n=3). Обработка данных экспрессии miRNA проводилась в программной среде R с применением пакета edgeR. Алгоритмами клеточной деконволюции выступали DeconRNASeq [2] (алгоритм, основанный на референсных значениях) и debCAM [3] (алгоритм, не требующий референса).

Результаты

Оценка дифференциально экспрессированных miRNA между различными клеточными линиями показала значительное сходство профилей экспрессии внутри клеточных линий ГМК, ЭК и Т-клеток. Помимо этого, ГМК и фибробласты, а также NK- и Т-клетки были объединены в общие группы («Т+НК» и «ГМК+ФБ»), по причине сходства их профилей экспрессии. Между объединёнными клеточными группами определялось от 177 (между Т- и В-клетками) до 428 (между «Т+НК» и «ГМК+ФБ») дифференциально экспрессируемых miRNA (в среднем 312). На основе медианного значения уровня экспрессии каждой

miRNA в каждой клеточной группе сформирован паттерн референсных сигналов для деконволюции, состоящий из 441 miRNA для 6 клеточных групп.

Точность метода оценивалась путем разложения первичного набора данных на клеточные компоненты с использованием библиотеки DeconRNASeq. Предварительная оценка показала, что из 40 образцов, 4 имели высокую долю «несвойственного» клеточного типа. Так, один образец ЭК был определён как «ГМК+ФБ» тип, другой образец ГМК имел 60% макрофагального компонента, а 2 образца «Т+НК» типа содержали ≈50% В-клеточного компонента. Мажорный компонент с долей >80% выявлялся для 27 образцов.

Для оценки применимости подходов деконволюции без референса был использован алгоритм debCAM с оптимальным разложением на 7 компонент. Все компоненты имели отражение в клеточных компонентах, определённых DeconRNASeq. Отличием было то, что тип «ГМК+ФБ» раскладывался на две компоненты. Вместе с тем, две компоненты характеризующие лимфоциты, позволяли выделить все 3 клеточных фенотипа, где Т-клетки занимали промежуточное значение между NK- и В-клетками. Дополнительной особенностью данного метода было то, что хотя для образцов определяемые клеточные доли были менее сконцентрированы в отдельных клеточных компонентах (только 9 образцов имели мажорный компонент с долей >80%), он позволяет на основании компонент классифицировать образцы со 100% точностью. Результаты исследования доступны по ссылке https://github.com/alekseizarubin/miRNA_reference_deconvolution_arteries

Разработанный алгоритм позволил получить эффективную референсную панель для деконволюции стенки артерии с использованием данных экспрессии miRNA. Алгоритм debCAM хорошо подходит для задачи классификации образцов и может работать без референса, обнаруживая новые клеточные субфенотипы. Алгоритм DeconRNASeq лучше решает задачу точного определения клеточных компонентов в образце.

Литература/ References

1. De Rie D. et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature biotechnology*. 2017;35(9):872-878.
2. Gong T., Szustakowski J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics*. 2013;29(8):1083-1085.
3. Newman A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015;12(5):453-457.