© Коллектив авторов, 2020

Булушева И.А.¹, Козлов И.Б.², Митин А.Н.², Коростин Д.О.¹, Кофиади И.А.²

Алгоритм анализа данных NGS при оценке репертуаров Т-клеточных рецепторов, вовлеченных в противоопухолевый иммунный ответ

¹ Центр высокоточного редактирования и генетических технологий для биомедицины Федерального государственного автономного образовательного учреждения высшего образования «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации, 117997, г. Москва, Российская Федерация

² Федеральное государственное бюджетное учреждение «Государственный научный центр «Институт иммунологии» Федерального медико-биологического агентства, 115522, г. Москва, Российская Федерация

Резюме

Введение. Репертуар Т- и В-клеточных рецепторов представлен большим числом последовательностей. В связи с этим характеристика иммунного рецепторного репертуара является сложной в техническом плане задачей. Наиболее современным методом изучения иммунных рецепторных репертуаров является NGS-секвенирование. Для снижения риска ошибки при работе с последовательностями используются методы вычислительной биологии, упрощающие обработку больших объемов данных.

Цель исследования – разработка и апробация алгоритма анализа данных NGS для сравнительной оценки репертуаров Т-клеточных рецепторов при сингенном и аллогенном ответе на опухоль.

Материал и методы. В рамках настоящей работы нами создана цепочка программных сценариев, позволяющих повысить эффективность анализа репертуара Т-клеточных рецепторов. Разработанные сценарии позволяют автоматически запускать обработку «сырых» данных в программном обеспечении MIGEC, объединять последовательности в пределах групп, проводить анализ клонального состава популяций Т-клеток с учетом специфичности иммунного ответа.

Для оценки сингенного ответа опухолевые клетки тимомы EL-4 прививали мышам линии C57BL/6, а для оценки аллогенного ответа – мышам линии B10.D2 (R101). Для исключения аллоспецифической составляющей при аллогенном ответе на опухоль дополнительная группа животных линии B10.D2 (R101) была иммунизирована нормальными тимоцитами мышей линии C57BL/6.

Результаты. При сравнении репертуаров Т-клеточных рецепторов в обследованных группах установлено, что противоопухолевый Т-клеточный ответ характеризуется снижением разнообразия и повышением клональности иммунного ответа на опухоль, наиболее выраженными в аллогенной группе мышей линии B10.D2 (R101). Анализ противоопухолевых клонов в сингенной и аллогенной модели продемонстрировал достоверные отличия качественного состава Т-клеточного репертуара. С помощью разработанных программных сценариев проведен анализ клонального состава Т-клеточных популяций с учетом аллоспецифической составляющей иммунного ответа и установлены конкретные клоны, обеспечивающие противоопухолевый ответ в сингенной и аллогенной модели.

Заключение. В результате исследования разработан алгоритм анализа данных NGS-секвенирования, прошедших обработку в программном обеспечении MIGEC. Условия и параметры анализа апробированы в биологическом эксперименте по сравнению противоопухолевого ответа в группах мышей, иммунизированных опухолевыми клетками линии EL-4. Проведено секвенирование α - и β -цепей Т-клеточных рецепторов у иммунизированных мышей и у контрольных животных. Проведен сравнительный анализ репертуаров Т-клеточных рецепторов у мышей в сингенной и аллогенной модели при развитии противоопухолевого иммунного ответа. Установлено, что сингенный и аллогенный противоопухолевый ответ, как и аллогенный ответ на интактную ткань, отличаются друг от друга. При этом в структуре иммунного ответа в аллогенной модели, по всей видимости, присутствуют опухоль-специфичные клоны, отсутствующие в сингенной модели.

Ключевые слова: Т-клеточный рецептор; NGS; оценка репертуара Т-клеточных рецепторов, противоопухолевый иммунный ответ

Для корреспонденции
Кофиади Илья Андреевич –
доктор биологических наук,
заведующий лабораторией
молекулярной иммуногенетики
ФГБУ «ГНЦ Институт иммунологии»
ФМБА России,
Москва, Российская Федерация
E-mail: kofiadi@mail.ru
<http://orcid.com/0000-0001-9280-8282>

Статья поступила 26.06.2020. Принята в печать 17.08.2020.

Для цитирования: Булусева И.А., Козлов И.Б., Митин А.Н., Коростин Д.О., Кофиади И.А. Алгоритм анализа данных NGS при оценке репертуаров Т-клеточных рецепторов, вовлеченных в противоопухолевый иммунный ответ. Иммунология. 2020; 41 (5): 400–410. DOI: <https://doi.org/10.33029/0206-4952-2020-41-5-400-410>

Финансирование. Исследование сингенного и аллогенного ответа на опухоль выполнено при финансовой поддержке Фонда перспективных исследований. Исследование по разработке алгоритма анализа данных выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-33-90076.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Bulusheva I.A.¹, Kozlov I.B.², Mitin A.N.², Korostin D.O.¹, Kofiadi I.A.²

NGS data analysis algorithm for evaluating repertoire of T-cell receptors involved in the antitumor immune response

¹ Center for High Precision Editing and Genetic Technologies for Biomedicine of N.I. Pirogov Russian National Research Medical University of Ministry of Health of the Russian Federation, 117997, Moscow, Russian Federation

² National Research Center Institute of Immunology of the Federal Medical and Biological Agency, 115522, Moscow, Russian Federation

Abstract

Introduction. The repertoire of T- and B cellular receptors is represented by a large number of sequences. Therefore, the characterization of the immune receptor repertoire is a technically difficult task. The most advanced method of analyzing the immune receptor repertoires is NGS sequencing. To reduce the risk of error when working with sequences, computational biology methods are used, which simplify the processing of large amounts of data.

Aim of study – development and testing of NGS data analysis algorithm for comparative evaluation of the T-cell receptor repertoire in syngeneic and allogeneic responses to the tumor.

Material and methods. As a part of this work we have created a chain of program scenarios that allow to increase the efficiency of analyzing the repertoire of T-cell receptors. The developed scenarios allow us to automatically start processing of «raw» data in MIGEC software, merge sequences within groups, and analyze clonal composition of T-cell populations taking into account the specificity of the immune response. The conditions and parameters of the analysis have been tested in a biological experiment in comparison with the antitumor response in groups of mice immunized by EL-4 line tumor cells.

Tumor cells of EL-4 line were grafted into C57BL/6 mice to evaluate the syngeneic response, and tumor cells of B10.D2 (R101) line were grafted into B10.D2 (R101) mice to evaluate the allogeneic response. To exclude allospecific component in allogeneic response to the tumor, an additional group of animals of B10.D2 (R101) line was immunized with normal thymocytes of C57BL/6 mice.

Results. When comparing the repertoire of T-cell receptors in the examined groups, it was found that the antitumor T-cell response is characterized by a decrease in diversity and an increase in the clonality of the immune response to the tumor, most pronounced in the allogeneic group of B10.D2 (R101) mice. The analysis of antitumor clones in the syngeneic and allogeneic models showed reliable differences in qualitative composition of T-cell repertoire. Using the developed program scenarios, the clonal composition of T-cell populations was analyzed, taking into account the allospecific component of the immune response, and specific clones providing antitumor response in the syngeneic and allogeneic models were identified.

Conclusion. As a result of the research, an algorithm for analyzing the data of NGS-sequencing processed in MIGEC software was created. The conditions and parameters of the analysis were tested in a biological experiment in comparison with the antitumor response in groups of mice immunized by EL-4 line tumor cells. T-cell receptor sequencing of α - and β -chains in immunized mice and control animals were performed. A comparative analysis of repertoire of T-cell receptors in mice in a syngeneic and allogeneic model in the development of antitumor immune response was carried out. It was found that the syngeneic and allogeneic antitumor response, as well as allogeneic response to intact tissue, differ from each other. The structure of the immune response in the allogeneic model seems to contain tumor-specific clones that are absent in the syngeneic model.

For correspondence

Ilya A. Kofiadi – Dr. Sci.,
Head of the Laboratory
of Molecular Immunogenetics
of NRC Institute of Immunology
FMBA of Russia,
Moscow, Russian Federation
E-mail: kofiadi@mail.ru
<http://orcid.com/0000-0001-9280-8282>

Keywords: T-cell receptor; NGS; evaluation of TCR repertoire; antitumor immune response

Received 26.06.2020. **Accepted** 17.08.2020.

For citation: Bulusheva I.A., Kozlov I.B., Mitin A.N., Korostin D.O., Kofiadi I.A. NGS data analysis algorithm for evaluating repertoire of T-cell receptors involved in the antitumor immune response. *Immunologiya*. 2020; 41 (5): 400–10. DOI: <https://doi.org/10.33029/0206-4952-2020-41-5-400-410> (in Russian)

Funding. The investigation of the syngeneic and allogeneic response to the tumor was carried out with financial support from the Perspective Research Foundation. Research to develop an algorithm for data analysis was carried out with financial support from RFBR under scientific project No. 19-33-90076.

Conflict of interests. The authors declare no conflict of interests.

Введение

Эффективность иммунного ответа зависит от предшествующего «антигенного опыта», физических и клинических параметров (пол, возраст, наличие иммунной патологии), а также генетических факторов. Вместе они формируют уникальный репертуар Т- и В-клеточных рецепторов у позвоночных, использующих RAG-зависимую рекомбинацию при формировании клеток адаптивного иммунитета [1].

В настоящее время данные о структуре и динамике иммунных репертуаров находят практическое применение в области иммунотерапии онкологических заболеваний [2, 3]. В частности, исследование иммунных репертуаров представляет интерес при разработке и оптимизации методик адаптивной иммунотерапии опухолей [4, 5]. В ряде исследований была показана применимость этого метода, однако вопрос повышения эффективности адаптивной иммунотерапии остается открытым [6]. Некоторые исследования указывают на усиление противоопухолевого иммунного ответа, связанное с развитием аллогенной иммунной реакции [7, 8]. Однако подробная структура Т-клеточного иммунного ответа в условиях развития такой реакции до сих пор не охарактеризована. Является ли аллогенный иммунный ответ одновременно и противоопухолевым? Если аллоспецифическая и противоопухолевая реакция опосредованы различными клоонами, можно ли провести четкую границу между ними? Можно ли использовать опухоль-специфичные клоны для повышения эффективности иммунотерапии? На сегодняшний день на эти вопросы нет ответов.

Во многом недостаточная изученность иммунного рецепторного репертуара при иммунотерапии связана с методическими сложностями при анализе. С развитием технологии высокоэффективного параллельного секвенирования (Next Generation Sequencing, NGS) методические ограничения во многом оказались сняты. Однако базовые технологии обработки данных NGS не позволяют полностью избавиться от ошибок, накопленных в ходе пробоподготовки и секвенирования, без потерь реального разнообразия. Для решения этой проблемы было разработано несколько алгоритмов, позволяющих повысить качество «сырых» данных [9–11]. Однако дальнейшие цели и задачи конкретных исследований могут сильно отличаться, поэтому универсального алгоритма работы с последовательностями на сегодняшний день нет. Не-

обходимость работы с большими объемами данных, объединение и исключение последовательностей в пределах сравниваемых групп, визуализация результатов анализа требуют привлечения значительного информационного ресурса и наличия базовых навыков программирования и тем самым ограничивают возможности практического применения метода NGS для характеристики иммунного рецепторного репертуара.

Наличие программного сценария, позволяющего оперировать «сырыми» и обработанными данными, значительно упрощает анализ. В связи с этим целью нашей работы стала разработка и апробация собственного алгоритма анализа данных NGS для оценки репертуаров Т-клеточных рецепторов (ТКР), работающего в комплексе с программой MIGEC. В результате исследования созданы системы уникальных названий образцов для работы с данными из разных запусков секвенатора и упорядочивания данных секвенирования. Также создан автоматизированный программный сценарий, позволяющий скачать, упорядочить и проанализировать данные по каждому образцу с помощью программы MIGEC и самостоятельно написанных программных сценариев. С помощью разработанных программных сценариев проведен сравнительный анализ репертуаров ТКР при развитии противоопухолевого ответа у мышей в сингенной и аллогенной модели.

Материал и методы

Экспериментальные животные. Исследование проводили на самках конвенциональных мышей линии C57BL/6, полученных из питомника «Столбовая» ФГБУН НЦБМТ ФМБА России, а также на конвенциональных мышках линии B10.D2 (R.101), любезно предоставленных экспериментально-биологической лабораторией ФГБУ «НМИЦ онкологии им Н.Н. Блохина» Минздрава России.

Культура опухолевых клеток. Клетки опухолевого штамма EL-4, использованные в настоящей работе, были получены из коллекции опухолевых штаммов, созданной в экспериментально-биологической лаборатории ФГБУ «НМИЦ онкологии им Н.Н. Блохина» Минздрава России.

Иммунизация экспериментальных животных. Для оценки сингенного иммунного ответа мышей линии C57BL/6 иммунизировали живыми клетками опухолевого штамма EL-4 внутривенно в дозе 0,1 млн

клеток на мышь. При достижении опухоли размера 1,0 см по одному из диаметров ее плотно перевязывали для предупреждения дальнейшего распространения. Мышей линий B10.D2 (R101) иммунизировали живыми клетками опухолевого штамма EL-4 или нормальными тимоцитами мыши линии C57BL/6 внутрибрюшинно в дозе 10,0 млн клеток на мышь. Через 2 мес сингенных и аллогенных животных реиммунизировали путем подкожного введения клеток опухолевого штамма EL-4 или нормальных тимоцитов мыши линии C57BL/6 в дозе 1,0 млн клеток на мышь. На 9-е сутки после реиммунизации животных забивали и извлекали селезенку.

Секвенирование. При подготовке образцов к секвенированию селезенку мышей освобождали от соединительной ткани, суспензировали в фосфатно-солевом буфере (ФСБ) с добавлением 0,1 % БСА и 2 мМ ЭДТА, удаляли эритроциты с помощью лизирующего буфера (eBioscience, США), дважды отмывали ФСБ. Численность клеток определяли путем подсчета в камере Горяева. Полученные на пике иммунного ответа спленоциты лизировали с выделением РНК методом фенол-хлороформной экстракции (реагент Trizol, Ambion, США). Библиотеки кДНК переменных фрагментов α - и β -цепей ТКР готовили в соответствии с протоколом, описанным ранее [12]. В итоге для анализа было собрано 40 образцов от 5 групп мышей:

- 1) «B6» – intactные мыши линии C57BL/6 (10 образцов);
- 2) «B6 + EL-4» – мыши линии C57BL/6, иммунизированные EL-4 (10 образцов);
- 3) «R101» – intactные мыши линии B10.D2 (R101) (10 образцов);
- 4) «R101 + EL-4» – мыши линии B10.D2 (R101), иммунизированные EL-4 (5 образцов);

Таблица 1. Образцы, включенные в исследование

Линия мышей	Иммунизация EL4	Иммунизация тимоцитами	Количество
C57BL/6	–	–	10
C57BL/6	+	–	10
R101	–	–	10
R101	+	–	5
R101	–	+	5

- 5) «R101 + Тц-B6» – мыши линии B10.D2 (R101), иммунизированные тимоцитами мышей линии C57BL/6 (5 образцов).

Всего было собрано 40 образцов (табл.1).

Для проведения секвенирования фрагментов CDR3 генов α - и β -цепей ТКР было подготовлено 80 библиотек кДНК (по 2 для каждого образца) в соответствии с протоколом, описанным ранее [11, 12]. Олигонуклеотиды для подготовки библиотек были любезно предоставлены лабораторией геномики адаптивного иммунного ответа Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН.

Качество приготовления библиотек кДНК определяли с помощью капиллярного гель-электрофореза на приборе Bioanalyzer 2100 (Agilent, США). Распределение фрагментов в подготовленных библиотеках соответствовало ожидаемому (рис. 1).

Секвенирование проводили на платформах Illumina MiSeq и Illumina HiSeq (США) в соответствии с инструкцией производителя. Оборудование для проведения секвенирования было предоставлено Центром высокоточного редактирования и генетических технологий для биомедицины ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России.

Рис. 1. Распределение длин фрагментов одной библиотеки репертуаров Т-клеточных рецепторов, полученное с помощью капиллярного гель-электрофореза на Agilent Bioanalyzer 2100

Основной пик имеет длину 640 п.н., что соответствует ожидаемому размеру.

Программное обеспечение и статистическая обработка. Для обработки «сырых» данных использовали программу MIGEC [10]. Программа позволяет кластеризовать последовательности в соответствии с уникальным молекулярным бар-кодом и удалить ошибки, накопленные в ходе секвенирования. Для оценки разнообразия ТКР использовали индекс разнообразия Шеннона. Расчет проводили по следующей формуле:

$$H = - \sum_{i=1}^S (P_i \cdot \ln P_i),$$

где H – индекс разнообразия Шеннона; $\sum$ – сумма от клона 1 до клона S ; S – количество клонов; P_i – доля конкретной последовательности в популяции. Для оценки достоверности отличий и вычисления p был использован U -критерий Манна–Уитни.

Результаты

Секвенирование образцов

Для оценки качества подготовки библиотек было проведено секвенирование на приборе Illumina MiSeq в режиме парноконцевого чтения с двумя бар-кодами. Результаты анализа представленности «сырых» данных по образцам приведены в табл. 2.

Данные секвенирования на платформе MiSeq были подвергнуты анализу в программе MIGEC. В результате доля клонотипов, прошедших встроенные в программу фильтры, составила от 81 до 94 %, что говорит о высоком качестве полученных библиотек.

Для получения полных данных по иммунным репертуарам было проведено 3 последовательных секвенирования полученных библиотек кДНК на приборе Illumina HiSeq в режиме парноконцевого прочтения. В среднем количество прочтений на каждую библиотеку составило 12 194 300 шт.

(минимум 2 095 225, максимум 31 637 095 шт.). Таким образом, были получены данные, достаточные для полной характеристики иммунных рецепторных репертуаров.

Разработка алгоритма анализа данных высокоэффективного параллельного секвенирования иммунных рецепторных репертуаров

В качестве хранилища данных, полученных в результате секвенирования, использовали облачное хранилище basespace. Для скачивания данных из хранилища было необходимо решить проблемы совпадения названий файлов из разных запусков секвенирования и неупорядоченности названия файлов. Также большое количество данных (960 конечных fastq.gz-файлов, рис. 2) потребовало создания программного сценария для автоматизации скачивания и «сшивания» данных для α - и β -цепей каждого из 40 образцов.

Для решения этих проблем были введены универсальные названия образцов формата «дата_номер_цепь» (например, «05092019_1_α»), был написан программный сценарий (язык Python), соотносящий между собой названия файлов, ссылки на данные хранилища basespace и названия образцов и создающий универсальную структуру данных, где в каждой строке указана информация по образцу, а по столбцам – дата, номер и уникальное название образца с 12 ссылками для скачивания последовательностей α - и β -цепей ТКР (по 6 для каждой цепи).

Для упорядочивания данных по образцам была создана единая архитектура хранения: корневая папка, совпадающая по названию с универсальным названием образца, содержащая папки для α - и β -цепей, включающие 3 папки (по одной для каждого из проведенных запусков секвенирования) (рис. 3).

Таблица 2. Количество прочтений и качество данных, полученных при секвенировании на платформе Illumina MiSeq

ID библиотеки	Количество необработанных прочтений	Количество необработанных прочтений, прошедших фильтрацию	Доля прочтений, прошедших фильтрацию, %
05092019_1_α	2847	2788	97,9
05092019_2_α	1545	1504	97,3
05092019_3_α	1465	1440	98,3
05092019_4_α	1694	1657	97,8
05092019_5_α	1468	1433	97,6
05092019_6_α	2481	2410	97,1
05092019_7_α	1098	1066	97,1
05092019_8_α	646	630	97,5
05092019_9_α	1457	1420	97,5
05092019_10_α	634	617	97,3
05092019_1_β	2628	2562	97,5
05092019_2_β	1590	1561	98,2
05092019_3_β	2225	2177	97,8
05092019_4_β	1480	1447	97,8
05092019_5_β	2052	2016	98,2
05092019_6_β	3204	3132	97,8
05092019_7_β	1275	1241	97,3
05092019_8_β	1239	1220	98,5
05092019_9_β	858	832	97,0
05092019_10_β	826	804	97,3
16012020_1_α	5136	5035	98,0

Окончание табл. 2

ID библио-теки	Количество необработанных прочтений	Количество необработанных прочтений, прошедших фильтрацию	Доля прочтений, прошедших фильтрацию, %
16012020_2_α	5617	5510	98,1
16012020_3_α	3942	3857	97,8
16012020_4_α	5401	5280	97,8
16012020_5_α	6776	6586	97,2
16012020_6_α	6839	6728	98,4
16012020_7_α	5387	5268	97,8
16012020_8_α	5057	4961	98,1
16012020_9_α	6083	5953	97,9
16012020_10_α	5763	5657	98,2
16012020_1_β	3724	3660	98,3
16012020_2_β	4643	4556	98,1
16012020_3_β	4173	4104	98,3
16012020_4_β	4574	4469	97,7
16012020_5_β	5957	5854	98,3
16012020_6_β	7124	6960	97,7
16012020_7_β	4769	4660	97,7
16012020_8_β	5032	4925	97,9
16012020_9_β	6562	6374	97,1
16012020_10_β	5190	5088	98,0
18112019_1_α	1229	1205	98,0
18112019_2_α	2252	2206	98,0
18112019_3_α	1963	1921	97,9
18112019_4_α	1722	1691	98,2
18112019_5_α	1943	1897	97,6
18112019_6_α	2576	2525	98,0
18112019_7_α	2582	2531	98,0
18112019_8_α	2381	2333	98,0
18112019_9_α	656	647	98,6
18112019_10_α	1155	1132	98,0
18112019_1_β	1249	1225	98,1
18112019_2_β	1794	1746	97,3
18112019_3_β	1939	1901	98,0
18112019_4_β	2157	2113	98,0
18112019_5_β	1785	1754	98,3
18112019_6_β	2865	2789	97,3
18112019_7_β	1356	1324	97,6
18112019_8_β	2280	2227	97,7
18112019_9_β	484	474	97,9
18112019_10_β	910	879	96,6
27112019_1_α	2017	1960	97,2
27112019_2_α	1809	1760	97,3
27112019_3_α	2067	2024	97,9
27112019_4_α	1585	1473	92,9
27112019_5_α	1340	1308	97,6
27112019_6_α	2215	2158	97,4
27112019_7_α	1849	1800	97,3
27112019_8_α	2474	2421	97,9
27112019_9_α	2386	2331	97,7
27112019_10_α	2357	2317	98,3
27112019_1_β	2254	2186	97,0
27112019_2_β	1851	1809	97,7
27112019_3_β	1532	1489	97,2
27112019_4_β	1982	1946	98,2
27112019_5_β	1090	1062	97,4
27112019_6_β	2039	1985	97,4
27112019_7_β	2441	2379	97,5
27112019_8_β	1932	1889	97,8
27112019_9_β	3240	3178	98,1
27112019_10_β	2038	2013	98,8

Рис. 2. Структура данных, полученных в результате трех независимых запусков секвенирования (run 1, run 2, run 3) для парноконцевых прочтений α - и β -цепей Т-клеточных рецепторов

Данная архитектура создается автоматически с помощью разработанного нами программного сценария в начале работы с образцом (язык Bash). Еще один сценарий (язык Bash) находит ссылки на хранилище basespace для соответствующего образца и скачивает последовательности в папки созданной архитектуры. Затем файлы сшиваются сначала в рамках одного запуска секвенирования, а затем для всех 3 запусков. В итоге в корневой папке образца формируются 2 файла (правое и левое прочтения) для α -цепи и β -цепи (всего 4 файла). Таким образом, разработанная совокупность программных сценариев позволяет объединить файлы без потери данных и сократить количество файлов с 24 для каждого образца до 4 файлов, необходимых для корректного иммунологического анализа.

На следующем этапе был написан программный сценарий, позволяющий автоматизировать процесс обработки «сырых» данных с помощью программы MIGEC. Его суть заключается в создании 2 txt-файлов в корневой папке образца, загружаемых в программу MIGEC, за-

Рис. 4. Архитектура хранения «сырых» и обработанных с помощью программы MIGEC данных

2 txt-файла, подаваемые на вход MIGEC отмечены стрелками. Запуск команд MIGEC, сопровождающихся созданием 5 папок (отмечены зеленым цветом). Необходимые «сырые» данные находятся в той же корневой папке (отмечены оранжевой рамкой).

Рис. 3. Пример архитектуры хранения данных для одного образца

пуске команд MIGEC, сопровождающемся созданием 5 папок. Необходимые «сырые» данные находятся в той же корневой папке. Итоговые обработанные данные загружаются в папку cdrfinal (рис. 4). Разработанные сценарии, объединенные в единый код, доступны для скачивания (https://github.com/genomecenter/TCR_article_2020).

Сравнительный анализ репертуаров Т-клеточных рецепторов у мышей в сингенной и аллогенной моделях при развитии противоопухолевого иммунного ответа

Разнообразие Т-клеточного ответа на опухоль EL-4

С помощью разработанного алгоритма была проведена оценка разнообразия репертуаров ТКР – количества уникальных клонотипов внутри образца. Для оценки ответа группы мышей использовали пулированные данные по группам. Объединение в данном случае подразумевает исключение повторяющихся клонов с формированием общей выборки, по сути, единого образца, характеризующегося нормированным по числу клонов разнообразием. Проводили сравнение для следующих групп:

- 1) 1+2 (сингенная модель)
- 2) 3+4+5 (иммунизация EL-4 и тимоцитами в аллогенной модели)

Результаты сравнения представлены на рис. 5.

Для оценки достоверности отличий использовали индекс разнообразия Шеннона (H). Этот индекс обычно используется для характеристики видового разнообразия в сообществе. Индекс Шеннона учитывает как численность, так и равномерность присутствующих видов. В случае с оценкой разнообразия за «вид» берется конкретный клонотип, S – общая совокупность клонотипов с учетом их долей в данном пулированном образце. По результатам анализа при одинаковом разнообразии для сравниваемых групп значение равномерности колебалось в диапазоне от 0,79 до 0,82, что говорит о наличии клонотипов, сильно выдающихся по количеству в общей массе, и о наличии большого числа клонотипов, присутствующих в незначительном количестве.

Для вычисления p был использован U -критерий Манна–Уитни. U -критерий Манна–Уитни позволяет сравнивать 2 группы данных, которые распределены не-

Рис. 5. Результаты оценки разнообразия α -цепей Т-клеточного рецептора в сингенной (А) и аллогенной модели (Б)

нормально. При оценке 20 000 наиболее представленных клонотипов при попарном сравнении групп 1+2, 2+4, 3+4, 4+5, 3+5 результаты, представленные в табл. 3.

Таким образом, все пулированные образцы достоверно отличаются между собой и процесс формирования набора клонотипов не является случайным.

Клональность Т-клеточного ответа на опухоль EL-4

Второй важный показатель, который был получен при оценке репертуара ТКР исследуемых групп – клональность – мера распространенности определенного клона. Разнообразие и клональность являются связанными характеристиками репертуаров. Высокое разнообразие всегда связано с низкой клональностью и наоборот.

По результатам оценки клональности Т-клеточного ответа было установлено, что модификация противоопухолевого ответа происходила за счет наиболее представленных клонов. Это наблюдение справедливо для всех сравниваемых групп интактных и иммунизированных мышей, однако наибольшее усиление клональности иммунного ответа происходило в аллогенной группе мышей линии B10.D2 (R101), иммунизированных EL-4, даже при сравнении с собственно аллогенной реакцией – ответом на интактные тимоциты мышей линии C57BL/6 (рис. 6).

Анализ 20 наиболее представленных клонов показал, что по качественным характеристикам репертуары в сингенной и аллогенной моделях практически не пересекаются (рис. 7).

Включение в исследование группы мышей, иммунизированных только аллогенными интактными тимоцитами, позволило нам провести анализ репертуаров на наличие опухоль-специфичных клонов. Для этого нами был разработан программный сценарий, позволяющий обнаружить и «вычистить» аллоспецифичные клоны из репертуара группы с аллогенным иммунным ответом на опухоль. В результате такого «вычитания» оставшиеся клоны потенциально являются опухоль-специфичными. В табл. 4 указаны последовательности наиболее представленных клонов, полученных после коррекции на аллоспецифический иммунный ответ.

Нами был проведен анализ этих последовательностей с помощью базы данных VDJdb и установлено, что 2 клон из выделенной нами группы – CASSPGQNTTEVFF и CASSLGGQNTLYF – были описаны ранее как противоопухолевые [13].

Обсуждение

В ходе проведенного исследования нами разработан программный алгоритм, позволяющий упростить и повысить эффективность анализа иммунного рецепторного репертуара. Создана единая архитектура и номенклатура образцов, разработан программный сценарий, создающий единую таблицу ссылок, разработаны 2 сценария для скачивания и комбинирования данных для каждого образца и создан сценарий, объединяющий все необходимые команды программы MIGEC. Данные программные сценарии объединены в единый код, запускаемый однократно и позволяющий сделать последовательный анализ от стадии «сырые данные» до стадии «полный обсчет образца» по введенным на входе 2 параметрам – дате и номеру образца. С помощью разработанного алгоритма был получен полный анализ для всех 40 образцов, включенных в исследование.

При анализе разнообразия в группах интактных и иммунизированных клетками опухолевой линии EL-4 мышей линии C57BL/6 установлена тенденция к уменьшению числа уникальных клонов у иммунизированных мышей (снижение разнообразия). Это может

Таблица 3. Результаты оценки 20 000 наиболее представленных клонотипов при попарном сравнении групп 1+2, 2+4, 3+4, 4+5, 3+5

Сравниваемые группы	<i>p</i>
1+2	0,00011128360323910641
2+4	2,6996821619839285e-23
3+4	2,8051445012743816e-39
4+5	2,8520770439120307e-17
3+5	1,0767775785964648e-07

Рис. 6. Сравнение клональности противоопухолевого иммунного ответа в группах intactных аллогенных мышей линии B10.D2 (R101), иммунизированных EL-4 (+EL-4) и нормальными тимоцитами мышей линии C57BL/6 (+Tц B6)

быть обусловлено относительным увеличением размеров клонов, вовлеченных в опухоль-специфический иммунный ответ.

При анализе разнообразия ТКР аллогенных по отношению к опухолевой линии EL-4 мышей линии B10.D2 (R101) [интактные B10.D2 (R101), иммунизированные EL-4, иммунизированные тимоцитами C57BL/6] обнаружены более значительные отличия по сравнению с сингенной группой. Разнообразие также снижено у иммунизированных мышей по сравнению с интактной группой, при этом у мышей, иммунизированных опухолевыми клетками (EL-4), в большей степени, чем у мышей, иммунизированных нормальными тимоцитами мышей линии C57BL/6.

Анализ 20 наиболее представленных клонов при сравнении аллогенного ответа на опухоль с сингенным противоопухолевым и аллоспецифическим от-

ветом на нормальные тимоциты продемонстрировал значительные отличия в качественном составе репертуаров ТКР у иммунизированных животных – последовательности практически не перекрывались. Таким образом, можно предполагать, что в репертуаре мышей B10.D2 (R101), иммунизированных EL-4, присутствовали уникальные клоны Т-клеток, которые обладали противоопухолевой специфичностью, не являлись аллореактивными и, таким образом, потенциально могут быть использованы при аллогенной адаптивной иммунотерапии.

Выполненная оценка репертуаров ТКР доказала целесообразность использования аллогенных опухоль-специфических Т-клеток для адаптивной иммунотерапии злокачественных новообразований при условии индукции толерантности к клеткам донора и удаления аллореактивных популяций из пула трансплантируемых клеток.

Рис. 7. 20 наиболее представленных клонов по экспрессии вариабельных участков α -цепи Т-клеточного рецептора (ТКР) при сингенном (B6) и аллогенном (R101) иммунном ответе на EL-4 (+EL-4) и аллогенном ответе на нормальные тимоциты мышей C57BL/6 (+Tц/B6)

Таблица 4. Нуклеотидные и аминокислотные последовательности CDR3-участков β-цепи Т-клеточного рецептора наиболее представленных клонов для выборки из группы аллогенных мышей B10.D2 (R101), иммунизированных EL-4

Количество	Частота	Нуклеотидная последовательность	Аминокислотная последовательность
105222	0,00551	TGTGCCAGCGGTGATGCAGGCAACTCCTATGAACAGTACTTC	CASGDAGNSYEQYF
86199	0,004514	TGTGCCAGCTCTCTCCATCTTGAAATACGCTCTATTTT	CASSLHSGNTLYF
81852	0,004287	TGTGCTAGCAGTCCCGGGACTGGGGGGTATGAACAGTACTTC	CASSPGTGGYEQYF
60405	0,003163	TGTGCCAGCAGCCCCGGGACAGTATGAACAGTACTTC	CASSPGQYEQYF
53803	0,002818	TGTGCCAGCAGTTCGGGGACTGGGGGGTATGAACAGTACTTC	CASSSGTGGYEQYF
49456	0,00259	TGTGCCAGCAGTTTAGGGACTGGGGGATATGAACAGTACTTC	CASSLGTGGYEQYF
49238	0,002579	TGTGCTAGCAGTTTAGGAGCAAACACAGAAGTCTTCTTT	CASSLGANTEVFF
46942	0,002458	TGTGCCAGCAGCCAAGCCAGGGGCTATGCTGAGCAGTTCTTC	CASSQARGYAEQFF
42195	0,00221	TGTGCCAGCAGCCAGGGGACTGGGTATGCTGAGCAGTTCTTC	CASSQGTGYAEQFF
42006	0,0022	TGTGCCAGCAGTCTGGGACAGTATGAACAGTACTTC	CASSLGQYEQYF
38061	0,001993	TGTGCCAGCAGCCCCGAGCTCCTATGAACAGTACTTC	CASSPSSYEQYF
37975	0,001989	TGTGCCAGCAGTCCGGGACAGAACACAGAAGTCTTCTTT	CASSPGQNTTEVFF
36880	0,001931	TGTGCCAGCAGTGGGACTGGAGGCTATGAACAGTACTTC	CASSGTGGYEQYF
35086	0,001837	TGTGCCAGCTCTCTCGGCGGGCAAACACCTTGTACTTT	CASSLGGQNTLYF
35014	0,001834	TGTGCCAGCAGCCCCGACAGTTCTGGAATACGCTCTATTTT	CASSPDSSGNTLYF
32910	0,001723	TGTGCCAGCTCTCTCGGACAAAACACAGAAGTCTTCTTT	CASSLGQNTTEVFF
31052	0,001626	TGTGCCAGCTCCCTCCAGGGGAACATGCTGAGCAGTTCTTC	CASSLQGNAYEQFF
30707	0,001608	TGTGCCAGCTCTCTCGGGGGTAACTATGCTGAGCAGTTCTTC	CASSLGGNYAEQFF
30302	0,001587	TGTGCCAGCTCTCTCGACAATTCTGAAATACGCTCTATTTT	CASSLDNSGNTLYF

Заключение

Созданный нами алгоритм анализа данных высокоэффективного параллельного секвенирования позволяет запускать обработку «сырых» данных в программном обеспечении MIGEC, объединять последовательности в пределах групп, производить анализ клонального состава популяций Т-клеток и оценивать отличия между сравниваемыми группами. Алгоритм апробирован в эксперименте по сравнению противоопухолевого ответа в группах мышей, иммунизированных опухолевыми клетками линии EL-4. Проведено секвенирование α- и β-цепей Т-клеточных рецепторов у иммунизированных мышей и у контрольных животных. Проведен сравнительный анализ репертуаров Т-клеточных рецепторов у мышей в сингенной и аллогенной моделях при развитии противоопухолевого иммунного ответа. Уста-

новлено, что сингенный и аллогенный противоопухолевый ответ, как и аллогенный ответ на интактную ткань, отличаются друг от друга. Удаление аллоспецифичных клонов из репертуара группы с аллогенным иммунным ответом на опухоль позволило определить клоны, отсутствующие в сингенной модели. Таким образом, иммунная реакция на опухоль развивается параллельно аллоспецифичной и может рассматриваться отдельно от нее. Соответствующие опухоль-специфичные клоны обладают потенциалом к использованию при иммунотерапии злокачественных новообразований.

Вклад авторов

Обработка методик, написание статьи – Булушева И.А., Козлов И.Б.; дизайн экспериментов – Митин А.Н.; секвенирование образцов – Коростин Д.О.; анализ данных, редактирование текста – Кофиади И.А.

Литература

1. Алексеев Л.П., Хаитов Р.М., Долбин А.Г., Болдырева М.Н., Алесеева П.Л., Трофимов Д.Ю. и др. Иммуногенетика человека и клиническая трансплантация органов в России. Иммунология. 2015; 36 (2): 76–89.
2. Han Y., Li H., Guan Y., Huang J. Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma. Cancer Lett. 2016; 379 36 (2): 206–12.
3. Ye B., Smerin D., Gao Q., Kang C., Xiong X. High-throughput sequencing of the immune repertoire in oncology: applications for clinical diagnosis, monitoring, and immunotherapies. Cancer Lett. 2018; 416: 42–56.
4. Bethune M.T., Joglekar A.V. Personalized T cell-mediated cancer immunotherapy: progress and challenges. Curr. Opin. Biotech. 2017; 48: 142–52.
5. Ho W.Y., Blattman J.N., Dossett M.L., Yee C., Greenberg P.D. Adoptive immunotherapy: engineering T cell responses as biologic weapons for tumor mass destruction. Cancer Cell. 2003; 3 (5): 431–7.
6. Rosenberg S.A., Restifo N.P., Yang J.C., Morgan R.A., Dudley M.E. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. Nat. Rev. Cancer. 2008; 8 (4): 299–308.
7. Kobayashi A., Hara H., Ohashi M., Nishimoto T., Yoshida K., Ohkohchi N. et al. Allogeneic MHC gene transfer enhances an effective antitumor immunity in the early period of autologous hematopoietic stem cell transplantation. Clin. Cancer Res. 2007; 13 (24): 7469–79.
8. Marcus A., Eshhar Z. Allogeneic adoptive cell transfer therapy as a potent universal treatment for cancer. Oncotarget. 2011; 2 (7): 525–6.
9. Bolotin D.A., Poslavsky S., Mitrophanov I., Shugay M., Mamedov I.Z., Putintseva E.V. et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods. 2015; 12 (5): 380–1.
10. Nazarov V.I., Pogorelyy M.V., Komech E.A., Zvyagin I.V., Bolotin D.A., Shugay M. et al. tcR: an R package for T cell receptor repertoire advanced data analysis. BMC Bioinformatics. 2015; 16 (1): 175.
11. Shugay M., Britanova O.V., Merzlyak E.M., Turchaninova M.A., Mamedov I.Z., Tuganbaev T.R. et al. Towards error-free profiling of immune repertoires. Nat. Methods. 2014; 11 (6): 653–5.
12. Mamedov I.Z., Britanova O.V., Zvyagin I.V., Turchaninova M.A., Bolotin D.A., Putintseva E.V. et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. Front. Immunol. 2013; 4: 456.
13. Dash P., Fiore-Gartland A.J., Hertz T., Wang G.C., Sharma S., Souquette A. et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature. 2017; 547 (7661): 89–93.

■ References

1. Alekseev L.P., Khaitov R.M., Dolbin A.G., Boldyreva M.N., Alekseeva P.L., Trofimov D.Yu., et al. Human immunogenetics and organ transplant clinics in Russia. *Immunologiya*. 2015; 36 (2): 76–89. (in Russian)
2. Han Y., Li H., Guan Y., Huang J. Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma. *Cancer Lett*. 2016; 376 (2): 206–12.
3. Ye B., Smerin D., Gao Q., Kang C., Xiong X. High-throughput sequencing of the immune repertoire in oncology: applications for clinical diagnosis, monitoring, and immunotherapies. *Cancer Lett*. 2018; 416: 42–56.
4. Bethune M.T., Joglekar A.V. Personalized T cell-mediated cancer immunotherapy: progress and challenges. *Curr. Opin. Biotech*. 2017; 48: 142–52.
5. Ho W.Y., Blattman J.N., Dossett M.L., Yee C., Greenberg P.D. Adoptive immunotherapy: engineering T cell responses as biologic weapons for tumor mass destruction. *Cancer Cell*. 2003; 3 (5): 431–7.
6. Rosenberg S.A., Restifo N.P., Yang J.C., Morgan R.A., Dudley M.E. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. *Nat. Rev. Cancer*. 2008; 8 (4): 299–308.
7. Kobayashi A., Hara H., Ohashi M., Nishimoto T., Yoshida K., Ohkohchi N., et al. Allogeneic MHC gene transfer enhances an effective antitumor immunity in the early period of autologous hematopoietic stem cell transplantation. *Clin. Cancer Res*. 2007; 13 (24): 7469–79.
8. Marcus A., Eshhar Z. Allogeneic adoptive cell transfer therapy as a potent universal treatment for cancer. *Oncotarget*. 2011; 2 (7): 525–6.
9. Bolotin D.A., Poslavsky S., Mitrophanov I., Shugay M., Mamedov I.Z., Putintseva E.V., et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods*. 2015; 12 (5): 380–1.
10. Nazarov V.I., Pogorelyy M.V., Komech E.A., Zvyagin I.V., Bolotin D.A., Shugay M., et al. tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics*. 2015; 16 (1): 175.
11. Shugay M., Britanova O.V., Merzlyak E.M., Turchaninova M.A., Mamedov I.Z., Tuganbaev T.R., et al. Towards error-free profiling of immune repertoires. *Nat. Methods*. 2014; 11 (6): 653–5.
12. Mamedov I.Z., Britanova O.V., Zvyagin I.V., Turchaninova M.A., Bolotin D.A., Putintseva E.V., et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front. Immunol*. 2013; 4: 456.
13. Dash P., Fiore-Gartland A.J., Hertz T., Wang G.C., Sharma S., Souquette A., et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. 2017; 547 (7661): 89–93.

Сведения об авторах

Булусева Ирина Алексеевна – аспирант Центра высокоточного редактирования и генетических технологий для биомедицины ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Москва, Российская Федерация; e-mail: bulusheva@phystech.edu, <http://orcid.org/0000-0001-8980-8425>

Козлов Иван Борисович – аспирант ФГБУ «ГНЦ Институт иммунологии» ФМБА России, Москва, Российская Федерация; e-mail: kozlov.nrcii@gmail.com, <http://orcid.org/0000-0002-1695-8737>

Митин Александр Николаевич – к.м.н., зав. лабораторией дифференцировки лимфоцитов ФГБУ «ГНЦ Институт иммунологии» ФМБА России, Москва, Российская Федерация; e-mail: mitin@inbox.ru, <http://orcid.org/0000-0003-1333-0757>

Коростин Дмитрий Олегович – руководитель группы секвенирования Центра высокоточного редактирования и генетических технологий для биомедицины ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Москва, Российская Федерация; e-mail: d.korostin@gmail.com, <http://orcid.org/0000-0003-1343-2550>

Кофиади Илья Андреевич – д.б.н., зав. лабораторией молекулярной иммуногенетики ФГБУ «ГНЦ Институт иммунологии» ФМБА России, Москва, Российская Федерация; e-mail: kofiadi@mail.ru, <http://orcid.org/0000-0001-9280-8282>

Authors' information

Irina A. Bulusheva – Postgraduate Student of the Center for High Precision Editing and Genetic Technologies for Biomedicine of the N.I. Pirogov RNRMU of the MOH of Russia, Moscow, Russian Federation; e-mail: bulusheva@phystech.edu, <http://orcid.org/0000-0001-8980-8425>

Ivan B. Kozlov – Postgraduate Student of NRC Institute of Immunology FMBA of Russia, Moscow, Russian Federation; e-mail: kozlov.nrcii@gmail.com, <http://orcid.org/0000-0002-1695-8737>

Alexander N. Mitin – PhD, Head of the Laboratory for Lymphocyte Differentiation of NRC Institute of Immunology FMBA of Russia, Moscow, Russian Federation; e-mail: mitin@inbox.ru, <http://orcid.org/0000-0003-1333-0757>

Dmitriy O. Korostin – Head of Sequencing Group of the Center for High Precision Editing and Genetic Technologies for Biomedicine of N.I. Pirogov RNRMU of the MOH of Russia, Moscow, Russian Federation; e-mail: d.korostin@gmail.com, <http://orcid.org/0000-0003-1343-2550>

Ilya A. Kofiadi – Dr.Sci., Head of Laboratory of Molecular Immunogenetics of NRC Institute of Immunology FMBA of Russia, Moscow, Russian Federation; e-mail: kofiadi@mail.ru, <http://orcid.org/0000-0001-9280-8282>