

РАЗРАБОТКА КАСТОМНЫХ БАРКОДОВ ДЛЯ СЕКВЕНИРОВАНИЯ НА ПЛАТФОРМЕ MGI

А. О. Шмитко , И. А. Булушева, Ю. А. Василиадис, О. Н. Сучалко, Д. С. Сырко, В. А. Белова, А. С. Павлова, Д. О. Коростин

Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва

Платформа секвенирования следующего поколения MGI (MGI Tech Co. Ltd., Китай), включающая секвенаторы DNBSEQ-G50, -G400 и -T7, активно внедряется в научные исследования. Однако сохраняются ограничения, связанные с использованием стандартных баркодов, в частности, на количество одновременно секвенируемых образцов и на соотношение их количества, а также присутствует проблема совместимости баркодов из разных или неполных наборов. Целью работы было разработать универсальный метод, позволяющий секвенировать до 252 образцов одновременно на одной дорожке секвенатора, с устранением ограничений, связанных с баркодами. Мы предложили «метод четверок», обеспечивающий уравнивание баркодов по принципу 4 или 4n+2. Проведена проверка метода на соответствие требованиям к баркодам для секвенирования, а также валидация метода на платформе DNBSEQ G-400. Метод четверок показал эффективность и надежность, позволяет секвенировать до 252 образцов одновременно без снижения качества данных, он оптимизирует процесс подготовки библиотек и повышает гибкость секвенирования на платформе MGI.

Ключевые слова: кастомные баркоды, кастомные индексы, кастомные адаптеры, DNBSEQ, MGI, BGI, NGS

Финансирование: соглашение о предоставлении из федерального бюджета грантов в форме субсидий в соответствии с пунктом 4 статьи 78.1 Бюджетного кодекса Российской Федерации на осуществление государственной поддержки создания и развития центра геномных исследований мирового уровня «Центр высокоточного редактирования и генетических технологий для биомедицины» № 075-15-2019-1789 от 22.11.2019.

Вклад авторов: А. О. Шмитко — планирование исследования, сбор данных, подготовка рукописи; И. А. Булушева — методология, анализ данных, подготовка рукописи; Ю. А. Василиадис, О. Н. Сучалко, Д. С. Сырко — анализ данных, программное обеспечение, визуализация; В. А. Белова — планирование исследования, рецензирование и редактирование рукописи; А. С. Павлова — курация данных, анализ данных, программное обеспечение; Д. О. Коростин — концептуализация, супервизия, методология, рецензирование и редактирование рукописи.

✉ **Для корреспонденции:** Анна Олеговна Шмитко
ул. Островитянова, д. 1, с. 1, г. Москва, 117997, Россия; annashmi97@gmail.com

Статья получена: 15.08.2024 **Статья принята к печати:** 17.09.2024 **Опубликована онлайн:** 10.10.2024

DOI: 10.24075/vrgmu.2024.040

DESIGNING OF CUSTOM BARCODES FOR SEQUENCING ON THE MGI PLATFORM

Shmitko AO , Bulusheva IA, Vasiliadis IA, Suchalko ON, Syrko DS, Belova VA, Pavlova AS, Korostin DO

Pirogov Russian National Research Medical University, Russia

The MGI (MGI Tech Co. Ltd., China) next-generation sequencing platform, including the DNBSEQ-G50, -G400, and -T7 sequencers, is being actively adopted in research. Despite its widespread adoption, challenges persist in the form of limitations associated with the manufacturer's provided barcode set for library preparation. These limitations include constraints on the number of samples that can be concurrently sequenced, compatibility issues with barcodes from diverse or incomplete sets, and restrictions on the sample ratio. Purpose: to develop a universal method that allows sequencing of up to 252 samples simultaneously on a single sequencer lane, while eliminating barcode-related limitations. We proposed a "quad method" that provides 4 or 4n+2 equilibration of barcodes. This paper also delves into its comprehensive analysis, verification procedures, seamless integration into the sequencing process and validation of the method on the DNBSEQ G-400 platform. The quad method showed efficiency and reliability, allowing sequencing of up to 252 samples simultaneously without compromising data quality. The proposed method optimizes library preparation and improves the flexibility of sequencing on the MGI platform.

Keywords: DNBSEQ, MGI, BGI, NGS, custom barcodes, custom indexes, custom adapters

Funding: this research was funded by the grant №075-15-2019-1789 from the Ministry of Science and Higher Education of the Russian Federation allocated to the Center for Precision Genome Editing and Genetic Technologies for Biomedicine.

Author contribution: Shmitko AO — study planning, data collection, writing Original Draft Preparation; Bulusheva IA — methodology, data analysis, writing Original Draft Preparation; Vasiliadis IA, Suchalko ON, Syrko DS — data analysis, software, visualization; Belova VA — study planning, manuscript review and editing; Pavlova AS — data curation, data analysis, software; Korostin DO — conceptualization, supervision, methodology, manuscript review and editing.

✉ **Correspondence should be addressed:** Anna O. Shmitko
Ostrovityanova, 1/1, Moscow, 117997, Russia; annashmi97@gmail.com

Received: 15.08.2024 **Accepted:** 17.09.2024 **Published online:** 10.10.2024

DOI: 10.24075/brsmu.2024.040

MGI Tech, основанная в 2016 г. как дочерняя компания BGI Group, является относительно новым игроком на рынке секвенирования следующего поколения (NGS) [1–3]. Первая платформа для секвенирования MGISEQ-2000 была представлена в 2017 г., за ней последовали платформы MGISEQ-200RS и MGISEQ-T7. MGI выпускает линейку секвенаторов на основе технологии наночастиц ДНК и секвенирования cPAS [4]. Секвенирование возможно в одноконцевом и парноконцевом режиме, с использованием одно- или двухбаркодных условий.

Технология предполагает баркодирование образцов на этапе лигирования адаптеров, содержащих их

последовательности. Баркодирование ДНК-библиотеки необходимо для маркирования последовательностей, принадлежащих разным биообразцам, а также для идентификации прочтений в процессе преобразования промежуточных файлов секвенирования в общепотребимый формат fastq. Длина последовательности баркода у MGI составляет 10 пар оснований.

Стандартные наборы реагентов для подготовки библиотек и секвенирования на приборе средней производительности DNBSEQ G-400 предназначены для одноиндексного секвенирования, тогда как режим двойного баркодирования является опциональным

и требует приобретения дополнительных наборов. В настоящее время MGI предлагает набор, включающий 96 баркодированных адаптеров для этапа лигирования при подготовке библиотеки ДНК для одноконцевого секвенирования. Кроме того, MGI предлагает 32 последовательности баркодов для синтеза.

Система G-400 чувствительна к балансу нуклеотидов на каждом цикле секвенирования баркода, так как качество резко падает, если одна и та же позиция в последовательностях баркодов с одной и той же дорожки занята одним и тем же нуклеотидом. По этой причине набор баркодов на одном лейне должен удовлетворять критериям комбинации их последовательностей и формировать совместимые сетки. Предлагаемый MGI набор из 96 адаптеров формирует 11 сетов (2 по 4, 8 по 8 и 1 на 24 баркода). Однако на практике возникает необходимость комбинировать образцы с баркодами из разных сетов, изменять количество образцов, загружаемых на дорожку ячейки, и их соотношение. В лабораторной практике нередко возникают ситуации, когда одна или несколько библиотек ДНК не соответствуют стандартам контроля качества на заключительном этапе. Гибкий подход к комбинированию образцов упрощает задачу объединения библиотек для загрузки на дорожку. Кроме того, необходимо учитывать задачу комбинирования образцов с разным требуемым объемом выходных данных, например экзоменов с разным покрытием $\times 200$ и $\times 100$.

Таким образом, производитель накладывает ограничения на пользователей платформы довольно небольшим набором баркодов и сетов, что не позволяет раскрыть ее истинный потенциал для секвенирования. Это может оказаться критичным при выборе платформы для секвенирования. Для Illumina разработаны индивидуальные решения для различных задач [5–7], в то время как для MGI таких решений пока нет.

Ранее мы разработали программное обеспечение, позволяющее выбрать оптимальную комбинацию предоставленных баркодов при различных соотношениях и количестве образцов для наборов адаптеров MGI [8]. Обновленное ПО, включающее пользовательские баркоды, доступно в репозитории GitHub (<https://github.com/genomecenter/BC-store/tree/custom-adapter-sets>). Для приборов Illumina для выбора сбалансированного соотношения баркодов в зависимости от задач секвенирования программы также существуют [9–11].

Цель данной работы — представить разработанный нами алгоритм, позволяющий генерировать необходимое для исследования количество последовательностей

баркодов. Используя его, мы подобрали 252 баркода, сформировав 63 сбалансированных сетов, каждый из которых состоит из 4 баркодов и позволяет комбинировать любые наборы сетов между собой.

МАТЕРИАЛЫ И МЕТОДЫ

Формулировка метода и подбор баркодов

Секвенатор имеет ограничения по интенсивности регистрируемого сигнала от флуорофоров, соответствующих нуклеотидам. Если в одной и той же позиции баркодов содержится одинаковый нуклеотид, то качество прочтения значительно снижается, что приводит к ошибкам идентификации баркодов и дальнейшего отнесения прочтений к образцам [8]. По этой причине перед нами стояла задача подобрать баркоды таким образом, чтобы они формировали наиболее сбалансированные комбинации. В основу алгоритма подбора последовательностей лег «метод четверок», который предполагает добавление к баркодам из набора MGI трех баркодов, полученных последовательной заменой оснований (рис. 1А, Б).

Согласно методу четверок, каждый из 96 баркодов может стать *root*-баркодом своей четверки. Таким образом можно получить $96 \times 4 = 384$ уникальных баркода.

Поскольку доля каждого основания в каждой позиции составляет 25%, полученная комбинация сбалансирована и гарантирует высокое качество секвенирования.

Проверка на соответствие критериям

Проверка на совместимость по сбалансированности

Поскольку каждая четверка является идеально сбалансированной, можно комбинировать любое количество четверок между собой. При этом соотношение долей четверок внутри пула может быть разным, однако внутри четверок доли баркодов должны быть одинаковыми.

Кроме того, мы проверили, возможно ли формировать пулы, содержащие $4n + 2$ баркода, где n — количество четверок. Проверку совместимости проводили на примере комбинации из 10 баркодов с помощью программы BC-store (рис. 2). Доля каждого нуклеотида (nucleotide fraction) в любой позиции в пуле из 10 баркодов имеет минимальное и максимальное отклонения, равные 0,2 и 0,3, и соответствует критериям сбалансированной комбинации. Это справедливо при добавлении двух

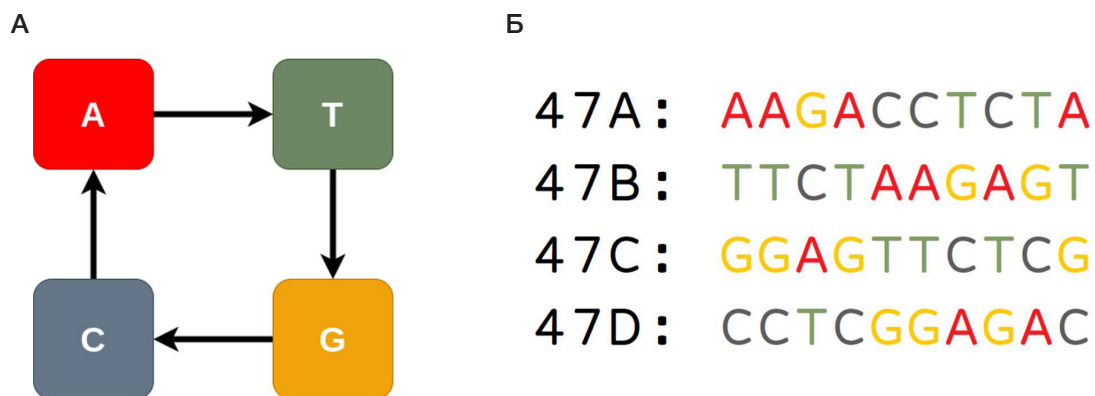


Рис. 1. А. Принцип метода четверок. Каждый баркод MGI является основой четверки, кастомные баркоды формируются с последовательной заменой в каждом положении исходного баркода А на Т, Т на G, G на C, C на А. Б. Пример четверки баркодов. 47А — исходный баркод MGI, 47В, 47С, 47D — сформированные на его основе по методу четверок кастомные баркоды

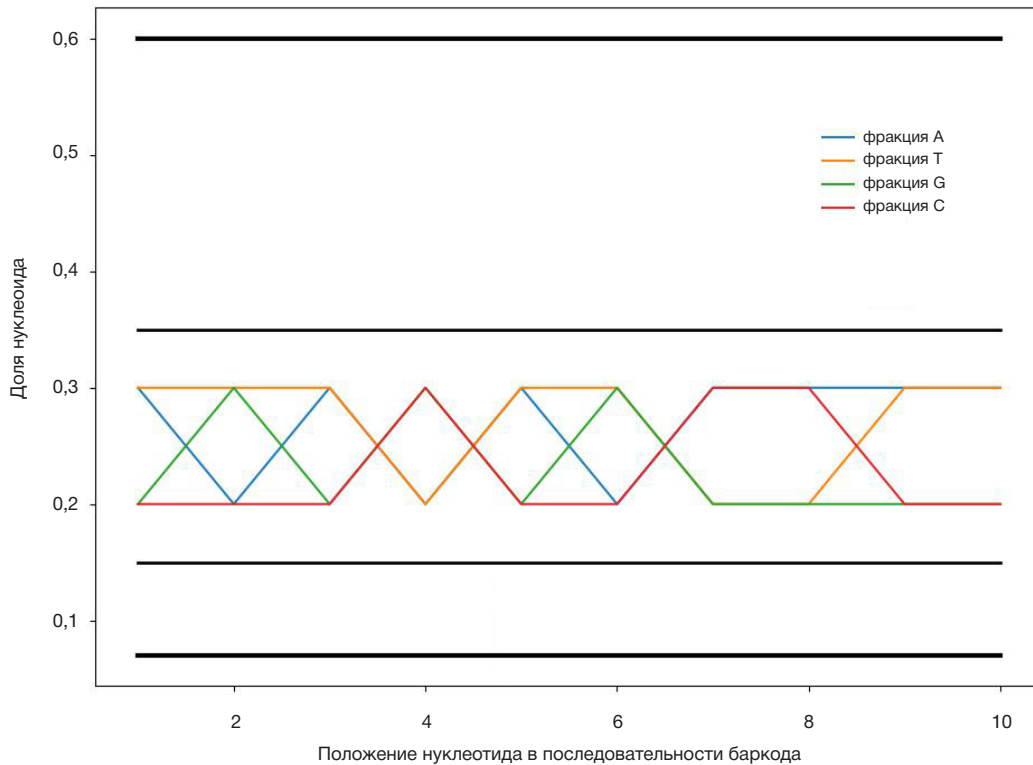


Рис. 2. Нуклеотидный баланс в пуле из $4n + 2$ баркода. Цветные линии представляют нуклеотидные фракции. Черные жирные линии — границы слабого критерия. Тонкие линии — границы сильного критерия совместимости баркодов [8]

любых баркодов из одной четверки в равном или меньшем соотношении к n количеству четверок.

Проверка на совместимость баркодов по количеству несовпадающих нуклеотидов

На следующем этапе все четверки были проверены на совместимость по количеству несовпадающих нуклеотидов (мисматчей). Каждый образец, маркированный баркодом, должен быть распознан однозначно, поэтому последовательности баркодов не должны совпадать на определенное количество нуклеотидов. Нами был выбран порог в 4 мисматча, так как все из 96 10-нуклеотидных баркодов производителя отличаются друг от друга не менее чем на четыре нуклеотида. В анализ также был включен баркод 999 верификации MGI (техническая последовательность длиной 10 п.н., присутствующая в файле демультиплексирования оригинального программного обеспечения). Построив граф несочетаемых четверок (см. приложение, рис. S1) и используя матрицу смежности (см. приложение, рис. S2), мы отобрали 63 четверки (252 баркода), совместимые друг с другом по количеству допустимых мисматчей (рис. 3). Последовательности всех 252 баркодов приведены в приложении (табл. S1).

Проверка на уникальность

Мы проверили, не встречаются ли предложенные нами последовательности баркодов среди оригинальных MGI. Это необходимо для составления файла с баркодами для автоматического демультиплексирования. Для этого была построена диаграмма Венна с набором кастомных и оригинальных MGI баркодов (рис. 4). В результате было получено 63 пересечения, где все 63 баркода — оригинальные MGI, и 189 — уникальные, не совпадающие с MGI из разных наборов.

Подготовка к секвенированию

Синтез адаптера

Согласно протоколу производителя, для создания индивидуального адаптера необходим отжиг двух олигонуклеотидов, один из которых (top) содержит последовательность баркода и фосфат на 5'-конце (Ad153_5T_1-index # (1~128) у производителя), а Bottom-олигонуклеотид имеет частично комплементарную верхнему (top) последовательность (Ad153Ω_Bottom_2) (https://en.mgitech.cn/Download/download_file/id/71) [12].

Последовательности олигонуклеотидов с баркодами 1A-1D приведены в табл. 1, полный список со всеми 252 баркодами в приложении (табл. S1).

Приготовление адаптеров осуществляли следующим образом. К 70 мкл LowTE буфера добавляли 1 мкл 5M NaCl, 10 мкл 200 мкМ top-олигонуклеотида, и 10 мкл 200 мкМ Bottom-олигонуклеотида. Далее нагревали до 95 °C в течение 2 мин и постепенно охлаждали до 17 °C с шагом 0,5 °C длительностью 30 с.

Алгоритм добавления новых баркодов на секвенатор

Для автоматического демультиплексирования секвенированных библиотек, следуя рекомендациям MGI, мы создали файл .csv (см. приложение, табл. S2), содержащий последовательности баркодов, включая новые кастомные, оригинальные MGI и 999-валидационный. Баркоды MGI, вошедшие в состав четверок, имели вид nA , где n — номер адаптера в оригинальном наборе MGI. Кастомные баркоды были обозначены nB , nC , nD — в соответствии с порядком формирования четверок. Запись оригинальных баркодов MGI, не вошедших в четверки, осталась неизменной. Номера баркодов были отделены от последовательностей запятыми без пробелов.

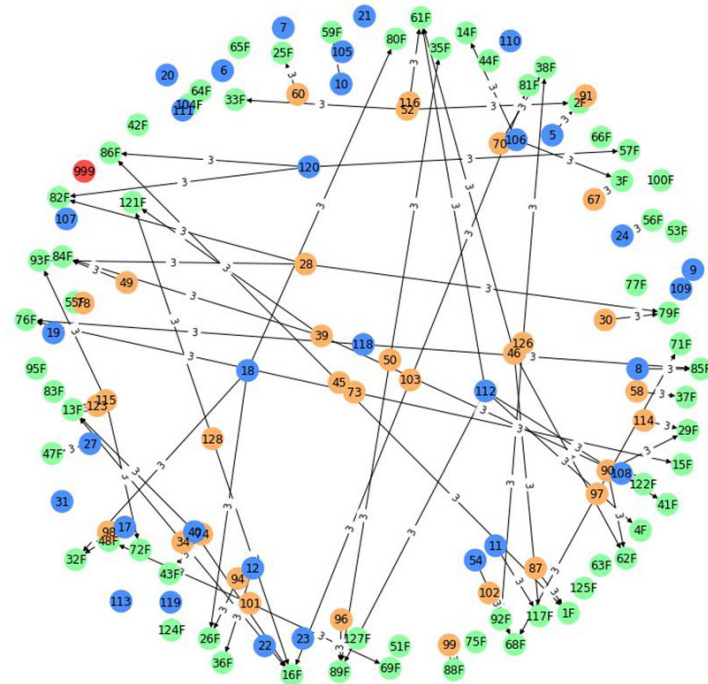


Рис. 3. Граф несочетаемости 63 четверок и баркодов MGI, не вошедших в их состав. Зеленым выделены четверки, прошедшие фильтрацию по числу мисматчей; оранжевым — баркоды из 96-баркодного набора MGI, не задействованные в четверках; синим — баркоды MGI из 128-баркодного набора MGI, отсутствующие в 96-баркодном наборе; красным — 999-проверочный баркод MGI. Линией соединены несовместимые баркоды и четверки, число над линией — минимальное значение мисматчей между ними

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для проверки разработанных нами баркодов, были подготовлены библиотеки с синтезированными кастомными адаптерами. Библиотеки, подготовленные по стандартному протоколу MGI, были объединены в пул и обогащены с помощью набора SureSelect Human All Exon v7 [13], а затем секвенированы в режиме PE100 на приборе DNBSEQ G-400.

Демультимплексирование Fastq выполнялось программным обеспечением, встроенным в G-400, MGI basecallite, на основе загруженного файла с последовательностями баркодов. По умолчанию алгоритм относит прочтение в категорию «undecoded» (не декодированные данные),

если в последовательности баркода длиной 10 п.н. есть два или более несоответствий. Таким образом доля недекодированных прочтений может быть использована в качестве метрики качества синтезированных адаптеров.

Мы сравнили долю недекодированных прочтений в общих данных с каждой дорожки с кастомными баркодами (44 дорожки ячеек для секвенирования) и данные предыдущих запусков (44 дорожки), в которых использовались баркоды MGI. В среднем (среднее ± SD) доля недекодированных чтений на одной дорожке составила $1,08 \pm 0,19\%$ и $1,68 \pm 0,22\%$ для адаптеров MGI и кастомных адаптеров соответственно (рис. 5). Хотя доля недекодированных данных увеличилась при использовании предложенных баркодов по сравнению с оригинальными

Сравнение наборов MGI и кастомного

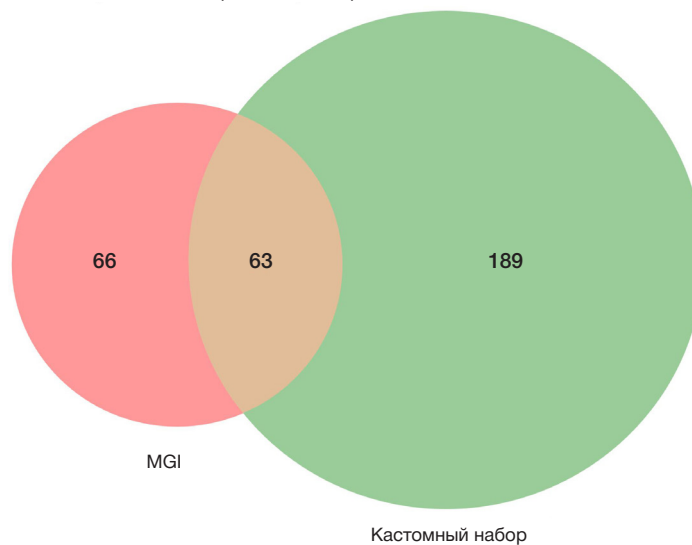


Рис. 4. Диаграмма Венна для сравнения последовательностей кастомных и оригинальных MGI баркодов из 128-баркодного набора и 999-го проверочного баркода MGI. Кастомные баркоды показаны зеленым цветом, оригинальные MGI, не пересекающиеся с четверками, — красным, а перекрывающиеся, которые были использованы в качестве root-баркодов для четверок, — оранжевым

Таблица. Пример последовательностей баркодов и полных последовательностей олигонуклеотидов для приготовления адаптеров. Ad_Bttm -bottom oligo

Имя	Последовательность баркода	Последовательность олигонуклеотида с баркодом
1A	TAGGTCCGAT	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAATAGGTCCGATCAACTCCTTGGCTCACA
1B	GTCCGAACTG	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAAGTCCGAACTGCAACTCCTTGGCTCACA
1C	CGAACTTAGC	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAACGAACCTTAGCCAACCTCCTTGGCTCACA
1D	ACTTAGGTCA	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAAACTTAGGTCAACTCCTTGGCTCACA
Ad_Bttm		TTGTCTTCTTAAGGAACGACATGGCTACGATCCGACTT

($T = 13.5$, $df = 83$, $p\text{-value} = 1.17E-22$), абсолютное значение по отношению к общему количеству данных, полученных с одной дорожки, можно считать незначительным. Отношение не декодированных данных к общим данным в ГБ представлены в приложении (табл. S3).

Таким образом, мы разработали подход для создания собственных баркодов, позволяющий одновременно секвенировать более 96 образцов на 1 дорожке секвенаторов MGI. Мы получили 189 кастомных баркодов, которые можно комбинировать с 63 баркодами MGI для создания 63 сбалансированных четверок. Один баркод из каждой четверки представляет собой оригинальный MGI (nA , где n — номер оригинального баркода), а три других — дополняющие его пользовательские баркоды (nB , nC , nD).

Эти четверки можно комбинировать друг с другом в любом количестве и соотношении, пока соотношение между баркодами одной четверки остается равным. Также возможно составлять пулы библиотек с числом баркодов $4n+2$, где n — число четверок, к которым можно добавить два любых баркода из другой четверки. В этом случае доли двух последних баркодов не должны превышать доли остальных.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Платформа MGI предназначена для быстрого высокопроизводительного секвенирования и обладает неоспоримыми преимуществами, но при этом имеет ряд ограничений. Мы попытались преодолеть некоторые ограничения, обусловленные решениями и наборами,

предлагаемыми производителем. Наш подход позволяет повысить эффективность секвенирования и расширить возможности платформы MGI. Однако важно учитывать, что комбинации четверок с некоторыми оригинальными адаптерами MGI, не входящими в состав четверок, могут не соответствовать критерию совместимости по числу несовпадений. Поэтому мы рекомендуем проверить их сбалансированность с помощью программы BC-Store. Мы предполагаем, что более высокое значение количества не декодированных чтений может быть связано с недостаточной чистотой синтезированных олигонуклеотидов по сравнению с MGI [14]. Ранее мы заказывали синтез одинаковых баркодов у двух разных производителей и наблюдали, что в случае одного из них доля не декодированных чтений была повышенной.

ВЫВОДЫ

Подобранные нами баркоды позволяют варьировать соотношение и количество библиотек, загружаемых на дорожку ячейки для секвенирования, в зависимости от задач и требуемого количества данных на выходе. В совокупности с применением разработанной нами ранее программы BC-store можно значительно упростить и ускорить задачу составления пулов библиотек для секвенирования на приборах MGI как в парноконцевом, так и одноконцевом режиме. Учитывая все преимущества и недостатки разработанного нами метода, его можно использовать в качестве дополнительного или альтернативного решения к предложенному производителем MGI.

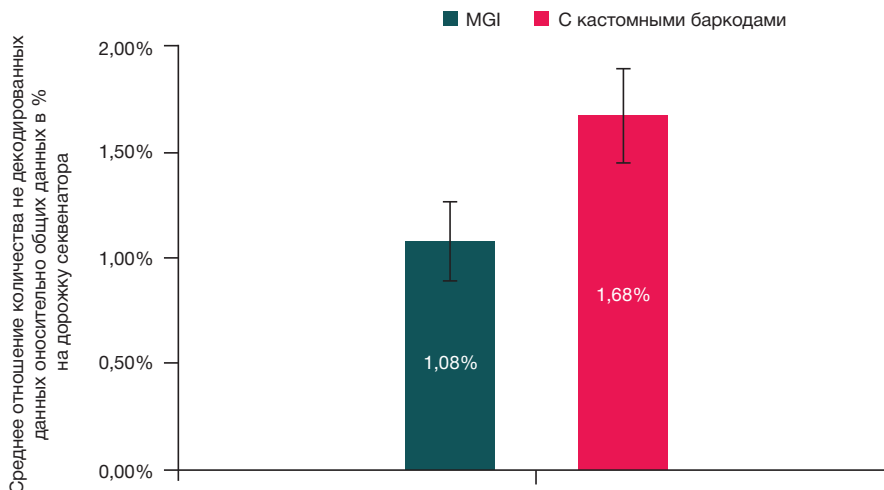


Рис. 5. Среднее значение доли undecoded (%) от общего числа полученных данных (Gb) на дорожку ячейки для библиотек с MGI-баркодами (показано голубым; данные для 44 дорожек), с кастомными баркодами (показано красным; данные для 44 дорожек)

Литература

1. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. *Genes Genomics*. 2021; 43: 713–24. DOI: 10.1007/s13258-021-01096-x.
2. Kim HM, Jeon S, Chung O, Jun JH, Kim HS, Blazyte A, et al. Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. *Gigascience*. 2021; 10. DOI: 10.1093/gigascience/giab014.
3. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. *BMC Bioinformatics*. 2019; 20. DOI: 10.1186/s12859-019-2751-3.
4. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* (1979). 2010; 327: 78–81. DOI: 10.1126/science.1181498.
5. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. 2014. Available from: <http://www.microbiomejournal.com/content/2/1/6>.
6. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*. 2015; 10. DOI: 10.1371/journal.pone.0128036.
7. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol*. 2013; 79: 5112–20. DOI: 10.1128/AEM.01043-13.
8. Bulusheva I, Belova V, Nikashin B, Korostin D. BC-store: A program for MGISEQ barcode sets analysis. *PLoS One*. 2021; 16. DOI: 10.1371/journal.pone.0247532.
9. Faircloth BC, Glenn TC. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS One*. 2012; 7. DOI: 10.1371/journal.pone.0042543.
10. Trébeau C, De Monvel JB, Tai FWJ, Petit C, Etournay R. DNA Barcode Compatibility: An R-package for optimizing DNA-barcode combinations in multiplex sequencing experiments. *Bioinformatics*. 2019; 35: 2690–1. DOI: 10.1093/bioinformatics/bty1030.
11. Varet H, Coppée JY. CheckMyIndex: A web-based R/Shiny interface for choosing compatible sequencing indexes. *Bioinformatics*. 2019; 35: 901–2. DOI: 10.1093/bioinformatics/bty706.
12. Oligos and primers for BGISEQ/DNBSEQ/MGISEQ - MGI Tech. Available from: https://en.mgitech.cn/Download/download_file/id/71.
13. SureSelectXT HS Target Enrichment System For Illumina Multiplexed Sequencing Platforms, Version D1, July 2021. Available from: <https://www.agilent.com/cs/library/usermanuals/public/G7530-90000.pdf>.
14. Li Q, Zhao X, Zhang W, Wang L, Wang J, Xu D, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. *BMC Genomics*. 2019; 20. DOI: 10.1186/s12864-019-5569-5.

References

1. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. *Genes Genomics*. 2021; 43: 713–24. DOI: 10.1007/s13258-021-01096-x.
2. Kim HM, Jeon S, Chung O, Jun JH, Kim HS, Blazyte A, et al. Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. *Gigascience*. 2021; 10. DOI: 10.1093/gigascience/giab014.
3. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. *BMC Bioinformatics*. 2019; 20. DOI: 10.1186/s12859-019-2751-3.
4. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* (1979). 2010; 327: 78–81. DOI: 10.1126/science.1181498.
5. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. 2014. Available from: <http://www.microbiomejournal.com/content/2/1/6>.
6. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*. 2015; 10. DOI: 10.1371/journal.pone.0128036.
7. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol*. 2013; 79: 5112–20. DOI: 10.1128/AEM.01043-13.
8. Bulusheva I, Belova V, Nikashin B, Korostin D. BC-store: A program for MGISEQ barcode sets analysis. *PLoS One*. 2021; 16. DOI: 10.1371/journal.pone.0247532.
9. Faircloth BC, Glenn TC. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS One*. 2012; 7. DOI: 10.1371/journal.pone.0042543.
10. Trébeau C, De Monvel JB, Tai FWJ, Petit C, Etournay R. DNA Barcode Compatibility: An R-package for optimizing DNA-barcode combinations in multiplex sequencing experiments. *Bioinformatics*. 2019; 35: 2690–1. DOI: 10.1093/bioinformatics/bty1030.
11. Varet H, Coppée JY. CheckMyIndex: A web-based R/Shiny interface for choosing compatible sequencing indexes. *Bioinformatics*. 2019; 35: 901–2. DOI: 10.1093/bioinformatics/bty706.
12. Oligos and primers for BGISEQ/DNBSEQ/MGISEQ - MGI Tech. Available from: https://en.mgitech.cn/Download/download_file/id/71.
13. SureSelectXT HS Target Enrichment System For Illumina Multiplexed Sequencing Platforms, Version D1, July 2021. Available from: <https://www.agilent.com/cs/library/usermanuals/public/G7530-90000.pdf>.
14. Li Q, Zhao X, Zhang W, Wang L, Wang J, Xu D, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. *BMC Genomics*. 2019; 20. DOI: 10.1186/s12864-019-5569-5.