

Артефакты анализа при идентификации клеточных линий методом профилирования на основе коротких tandemных повторов

Мальченкова А. А., Кособокова Е. Н.

ФГБУ "Национальный медицинский исследовательский центр онкологии им. Н. Н. Блохина" Минздрава России. Москва, Россия

Цель. Изучить и описать самые распространенные виды артефактов, которые возникают при детекции коротких tandemных повторов STR (short tandem repeat)-ампликонов методом капиллярного электрофореза и вызывают трудности в интерпретации полученных STR-профилей.

Материал и методы. Клеточные линии получены из биоресурсной коллекции (БРК) клеточных линий ФГБУ "НМИЦ онкологии им. Н. Н. Блохина" Минздрава России. Выделение ДНК из суспензий клеток проводили по инструкциям производителей наборов DNeasy Blood & Tissue (QIAGEN, Германия) и ExtractDNA Blood & Cells (Eurogen, Россия). Измерение концентрации ДНК осуществляли с помощью прибора Qubit 4.0 (Thermo Fisher Scientific, США) и набора реагентов Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, США). Мультиплексная полимеразная цепная реакция (ПЦР) была поставлена с помощью набора реактивов CoRDIS "ЭКСПЕРТ26" (Гордиз, Россия). Капиллярный электрофорез продуктов ПЦР проводили на генетическом анализаторе Genetic Analyzer 3500xL (Applied Biosystems, США). Для обработки электрофореграмм использовали программу GeneMapper Software v6.0 (Thermo Fisher Scientific, США).

Результаты. Изучены наиболее известные артефакты, связанные с методом STR-профилирования и последующим капиллярным электрофоретическим разделением ампликонов. Приведены случаи детектирования этих артефактов из личной практики. Даны рекомендации действий для улучшения картины электрофореза.

Заключение. В работе изучены артефакты анализа при идентификации результатов STR-профилирования клеточных линий методом капиллярного электрофореза (STR-CE), с которыми исследователи сталкиваются в лабораторной практике. Подробно описаны

и проиллюстрированы примерами из собственной практики распространенные виды артефактов анализа, которые вызывают трудности в интерпретации полученных при STR-профилировании результатов, а также возможные причины их появления. Приведены рекомендации по снижению количества неспецифических флуоресцентных сигналов и их интенсивности.

Ключевые слова: STR-профилирование, артефакты анализа, капиллярный электрофорез.

Отношения и деятельность. Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках научно-исследовательской работы № 075-15-2021-1060 "Создание и развитие биоресурсной коллекции генетически и фенотипически охарактеризованных клеточных линий и первичных опухолей человека".

Поступила 24/07-2024

Рецензия получена 15/09-2024

Принята к публикации 30/09-2024



Для цитирования: Мальченкова А. А., Кособокова Е. Н. Артефакты анализа при идентификации клеточных линий методом профилирования на основе коротких tandemных повторов. *Кардиоваскулярная терапия и профилактика*. 2024;23(11):4121. doi: 10.15829/1728-8800-2024-4121. EDN MNQAWT

Artifacts of analysis in cell line identification by short tandem repeat profiling

Malchenkova A. A., Kosobokova E. N.

Blokhin National Medical Research Center of Oncology. Moscow, Russia

Aim. To study and describe the most common types of artifacts in detection of short tandem repeat (STR) amplicons by capillary electrophoresis and cause difficulties in interpreting the obtained STR profiles.

Material and methods. Cell lines were obtained from the bioresource collection of cell lines of the Blokhin National Medical Research Center of Oncology. DNA was isolated according to the manufacturer's instructions of the DNeasy Blood & Tissue (QIAGEN, Germany) and ExtractDNA Blood & Cells (Eurogen, Russia) kits. DNA concentration

was measured using a Qubit 4.0 device (Thermo Fisher Scientific, USA) and a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, USA). Multiplex PCR was performed using a CoRDIS EXPERT26 reagent kit (Gordiz, Russia). Capillary electrophoresis of PCR products was performed on a 3500xL Genetic Analyzer (Applied Biosystems, USA). GeneMapper Software v6.0 (Thermo Fisher Scientific, USA) was used to process electrophoresis data.

Results. The most well-known artifacts associated with the STR profiling and subsequent capillary electrophoretic separation of

*Автор, ответственный за переписку (Corresponding author):

e-mail: ekkos@mail.ru

[Мальченкова А. А. — м.н.с. молекулярно-генетической диагностики и персонализированной медицины НИИ экспериментальной диагностики и терапии опухолей, ORCID: 0000-0003-2201-0472, Кособокова Е. Н. — в.н.с., к.б.н. лаборатории молекулярно-генетической диагностики и персонализированной медицины НИИ экспериментальной диагностики и терапии опухолей, ORCID: 0000-0002-4660-8519].

amplicons were studied. Cases of detection of these artifacts from personal practice are given. Recommendations for improving the electrophoresis pattern are given.

Conclusion. The paper studies the artifacts of analysis in cell line STR profiling by capillary electrophoresis (STR-CE), which researchers encounter in laboratory practice. Common types of analysis artifacts that cause difficulties in interpreting the results obtained during STR profiling, as well as possible reasons for their occurrence, are described in detail and illustrated with examples from our own practice. Recommendations are given for reducing the number of non-specific fluorescent signals and their intensity.

Keywords: STR profiling, analysis artifacts, capillary electrophoresis.

Relationships and Activities. The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation within the research work № 075-15-2021-1060 "Creation

and development of a bioresource collection of genetically and phenotypically characterized cell lines and primary human tumors".

Malchenkova A. A. ORCID: 0000-0003-2201-0472, Kosobokova E. N. * ORCID: 0000-0002-4660-8519.

*Corresponding author: ekkos@mail.ru

Received: 24/07-2024

Revision Received: 15/09-2024

Accepted: 30/09-2024

For citation: Malchenkova A. A., Kosobokova E. N. Artifacts of analysis in cell line identification by short tandem repeat profiling. *Cardiovascular Therapy and Prevention*. 2024;23(11):4121. doi: 10.15829/1728-8800-2024-4121. EDN MNQAWT

ДНК — дезоксирибонуклеиновая кислота, ПО — программное обеспечение, ПЦР — полимеразная цепная реакция, LOD — Limit of Detection (предел обнаружения), LOQ — Limit of Quantification (предел количественного определения), NGS — Next Generation Sequencing (секвенирование нового поколения, высокопроизводительное секвенирование), RFU — relative fluorescent unit (относительная единица флуоресценции), STR — short tandem repeat (короткие tandemные повторы), STR-CE — short tandem repeat, products are separated by size using capillary electrophoresis (профилирование на основе коротких tandemных повторов с последующим анализом ампликонов методом капиллярного электрофоретического разделения).

Ключевые моменты

Что известно о предмете исследования?

- Метод профилирования на основе коротких tandemных повторов (STR — short tandem repeat) является одним из самых простых и удобных методов для подтверждения подлинности клеточных линий, однако интерпретация результатов может быть осложнена артефактами анализа.

Что добавляют результаты исследования?

- В работе впервые на русском языке описаны распространенные артефакты анализа, которые были обнаружены в ходе экспериментальной работы методом STR-профилирования, а также возможные причины их возникновения. Кроме того, даны рекомендации для снижения степени влияния упомянутых артефактов на результаты анализа.

Key messages

What is already known about the subject?

- The short tandem repeat (STR) profiling is one of the simplest and most convenient methods for cell line authentication, but interpretation of the results can be complicated by analysis artifacts.

What might this study add?

- The work describes for the first time in Russian the common analysis artifacts discovered during the experimental work using the STR profiling, as well as the possible reasons for their occurrence. In addition, recommendations are given for reducing the influence of the mentioned artifacts on the analysis results.

Введение

Идентификация клеточных линий при проведении доклинических исследований, даже спустя 50 лет после обнаружения научным сообществом проблемы ошибочной идентификации и невозпроизводимости экспериментов, всё еще не является повсеместной рутинной лабораторной практикой [1-4]. Несмотря на то, что заявленная проблема довольно хорошо освещается в научном сообществе (база данных ошибочно идентифицированных клеточных линий ICLAC пополняется каждый год), выходят обзорные статьи, подсчитывающие экономический и репутационный урон, возникающий из-за недостаточного внимания к проблеме идентификации и развенчанных исследований, построенных на ошибочных гипотезах.

При изучении внушительного набора статей, описывающих проблему внутривидовой контамина-

ции и связанных последствий [5-9], становится очевидна необходимость повсеместного внедрения процедур идентификации биологического материала.

Актуальность настоящей статьи обусловлена, с одной стороны, тем, что в российском научном сообществе за счет увеличения числа международных публикаций растет внимание к проблеме контаминации клеточных линий в лабораториях и осознание необходимости устанавливать подлинность своих культур, а с другой стороны — отсутствием достаточного количества литературы на русском языке для погружения в тему.

Одним из доступных и удобных вариантов идентификации клеточных линий человека является STR (short tandem repeat)-профилирование [10], которое заключается в проведении мультиплексной полимеразной цепной реакции (ПЦР) с последующим разделением и анализом продуктов ампли-

фикации с помощью капиллярного электрофореза. Однако из-за артефактов, обусловленных физико-химическими особенностями используемых методов (ПЦР, электрофоретическое разделение, флуоресцентная детекция), результаты анализа не всегда просто интерпретировать.

Для идентификации клеточных линий в рамках исследовательских лабораторий на территории нашей страны STR-профилирование только входит в рутинную практику и ещё недостаточно широко распространено. Методологической литературы, кроме руководств производителей для наборов реагентов, на русском языке нет. Поэтому настоящая статья может быть полезна исследователям в интерпретации данных применительно к задаче генотипирования образцов клеточных линий человека.

Цель работы — изучить и описать самые распространенные виды артефактов, которые возникают при детекции STR-ампликонов методом капиллярного электрофореза и вызывают трудности в интерпретации полученных STR-профилей.

Материал и методы

В работе использовали клеточные линии из биоресурсной коллекции (БРК) ФГБУ "НМИЦ онкологии им. Н. Н. Блохина" Минздрава России. Выделение дезоксирибонуклеиновой кислоты (ДНК) из суспензий клеток проводили по инструкциям производителей наборов DNeasy Blood & Tissue (QIAGEN, Германия) и ExtractDNA Blood & Cells (Евроген, Россия). Концентрацию ДНК измеряли с помощью прибора Qubit 4.0 (Thermo Fisher Scientific, США) и набора реагентов Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, США). Мультиплексная ПЦР была поставлена с помощью набора реактивов CoRDIS "ЭКСПЕРТ 26" (Гордиз, Россия). Капиллярный электрофорез продуктов ПЦР проводили на генетическом анализаторе Genetic Analyzer 3500xL (Applied Biosystems, США). Для обработки электрофореграмм использовали программу GeneMapper Software v6.0 (Thermo Fisher Scientific, США).

Результаты и обсуждение

Биобанкам как хранителям и распространителям биологического материала необходимо осуществлять идентификацию и строгий контроль клеточных линий для повсеместного искоренения проблемы контаминации клеточных культур [11, 12].

Поэтому в рамках проекта по созданию БРК на базе НМИЦ онкологии им. Н. Н. Блохина был внедрен целый набор процедур по контролю качества и подлинности депонированных и получаемых клеточных линий. В качестве основного метода для идентификации клеточных линий, подтверждения их подлинности и чистоты был выбран метод STR-профилирования [13].

Данный метод отличается сочетанием относительной простоты анализа и надежностью результатов в отличие от традиционных методов (HLA-

типирования, кариотипирования и др.); его стоимость значительно ниже проведения NGS (Next Generation Sequencing, секвенирование нового поколения), а кроме того, начиная с 2022г, метод рекомендован в Национальном руководстве по биобанкированию [14]. К несомненным преимуществам данной методики можно отнести и наличие внушительного количества анонсированных в базах данных референсных STR-профилей¹ для наиболее распространенных клеточных культур, что позволяет контролировать воспроизводимость результатов, полученных в лаборатории. Однако стоит признать, что у данной методики, несмотря на то, что она, согласно международному стандарту по аутентификации клеточных линий человека, разработанному ATCC (The American Type Culture Collection — Американская коллекция типовых культур), считается "золотым стандартом" и широко используется, есть ограничения и сложности применения [15].

К ограничениям метода стоит отнести невозможность обнаружения межвидовой контаминации, поскольку представленные на рынке наборы реагентов содержат праймеры, высокоспецифичные к человеческой ДНК, что не позволяет перекрыть одним анализом все проблемы с контаминацией в лаборатории. Хотя для персонализированной идентификации некоторых животных уже разработаны STR-праймеры (мыши, крысы, обезьяны, некоторые виды рыб) [16–18] и разработка стандарта по STR-идентификации мышинных клеточных линий находится на заключительных стадиях, коммерчески доступных наборов реагентов для STR-профилирования лабораторных животных всё ещё не представлено. Некоторые биобанки (например, ATCC — American Type Culture Collection) и коммерческие фирмы, (ООО "Гордиз", АО "SynGen") предлагают STR-профилирование мышинных линий только в качестве услуги [19].

Ещё одним ограничением метода является невозможность определить точную нуклеотидную последовательность амплифицированных фрагментов [20]. Это означает, что детектируемые в локусе продукты одинакового молекулярного веса могут быть разными по нуклеотидному составу, но на электрофореграмме они между собой никак различаться не будут. Это приводит к тому, что, при обнаружении нетипичной электрофоретической картины для клеток с известным STR-профилем (выпадение аллеля, гетерозиготный дисбаланс аллелей в локусе, изменения количества tandemных повторов в аллеле и т.д.), определить точную причину явления удается не всегда [15].

Причиной детектируемых генетических изменений могут быть как точечные мутации в зоне отжига праймеров, так и кардинальные изменения

¹ <https://iclac.org/databases/cross-contaminations/>.

в самом генотипе клеток данной линии, возникновение субклональной популяции, а также контаминация другой клеточной линией или ошибочная идентификация клеток персоналом. Для выяснения генетической картины и установления первопричины необходимы исследования дополнительными методами (например, кариотипирование, секвенирование экзона).

Конечно, наборы реагентов, условия постановки STR-профилирования и процедура анализа данных совершенствуются и уже не представляют тот объем проблем, с которым исследователи сталкивались 20 лет назад [21]. И если раньше продукты амплификации приходилось разделять с помощью горизонтального электрофореза, что представляло собой трудоёмкую задачу и сложную в интерпретации электрофоретическую картину [22], то на сегодняшний день большинство лабораторий использует приборы для капиллярного электрофоретического разделения продуктов мультиплексной амплификации с флуоресцентной детекцией. Это существенно упрощает и ускоряет анализ, позволяет очень точно различать аллели по молекулярному весу, а также получить и обработать результаты в удобной электронной форме. Кроме того, приборы с несколькими каналами флуоресцентной детекции позволяют усложнить мультиплексную смесь и использовать большее количество генетических маркеров в одном исследовании.

Однако при всех неоспоримых преимуществах данной технологии, у приборов капиллярного электрофореза есть два основных ограничения, связанных с флуоресцентной детекцией. Первое ограничение — это наличие одного лазера (канала возбуждения) и нескольких каналов детекции в приборах. Для работы по нескольким каналам детекции в ПЦР-смесях содержится смесь флуорофоров. Сигналы от флуорофоров с близкими по значению длинами волн испускания флуоресценции частично перекрываются, например, сигнал от синего флуорофора частично попадает и регистрируется в диапазоне зеленого канала детекции. Такая "засветка" в соседний канал уменьшает относительную яркость истинного сигнала от флуорофора, и при нормировании прибором сигнала в некоторых каналах детекции теряется до 99% сигнала, что может приводить к значимым проблемам с детекцией результатов [23, 24].

Аналитические алгоритмы прибора, как и некоторые технологические решения, которые применяют производители наборов реагентов, частично сглаживают данный эффект, но иногда это не решает проблему полностью, а оператор наблюдает её проявление в виде артефактов на электрофореграмме.

Артефакты — это вторая широко распространенная проблема, возникающая у оператора при

анализе данных STR-исследования методом STR-CE. Именно об артефактах далее пойдет речь в этой статье.

Артефактами в STR-CE являются все сигналы, наблюдаемые на электрофореграмме, нетипичные для характерной картины электрофореза образцов ДНК человека.

Можно выделить основные причины появления артефактов на электрофореграммах:

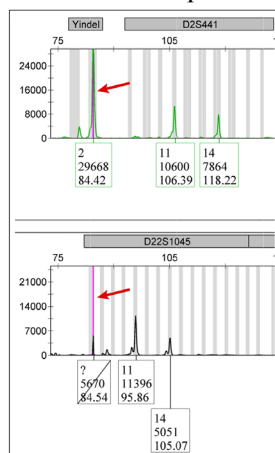
- качество и состояние реагентов для постановки анализа;
- качество загружаемых в прибор продуктов амплификации;
- неоптимальные настройки генетического анализатора (напряжение, время инъекции и др.);
- физические причины (стохастические эффекты, эффекты просачивания красителя);
- генетические особенности объекта исследования.

И если для преодоления первого ограничения методики, связанного с технической составляющей прибора, оператор ничего не может предпринять, то с артефактами анализа и алгоритмами уменьшения их количества вполне можно работать в рамках метода. После формирования ДНК-профилей тем или иным программным обеспечением (ПО) для анализа оператор должен проанализировать графики и оценить качество данных, отличить истинные сигналы от артефактов и, в конце концов, объединив данные от всех локусов, сформировать общий STR-профиль образца.

При продолжительной работе с методом и анализе данных STR-профилирования мы сталкивались с проявлениями всех основных артефактов анализа (рисунок 1), которые затрудняли интерпретацию результатов [13]. Некоторые артефакты не распознавались ПО прибора (в нашем случае GeneMapper Software 6.0), другие — приводили к ошибочным заключениям и формированию ПО некорректных STR-профилей, поэтому важно было проверять результаты вручную.

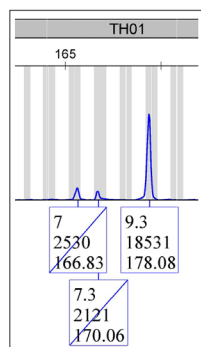
Поскольку криминалисты по всему миру уже около 30 лет [25, 26] используют STR-профилирование в своей повседневной практике для идентификации личности человека, для них разработано достаточное количество руководств и пособий по интерпретации сложных случаев из следственной практики. Также разработаны некоторые алгоритмы принятия решений о природе сигнала, основанные на разных математических подходах [27, 28]. Однако специфика исследуемых в биологических лабораториях клеточных культур и биобанках не позволяет применять на практике многие рекомендации, составленные для криминалистов. Генетическая интерпретация STR-профилей клеточных линий во многом отличается от интерпретации криминалистических образцов, потому что из-

Подтягивание красителя

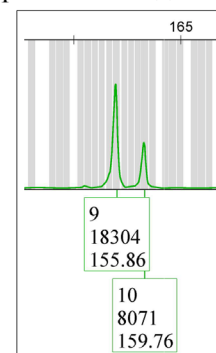


Виды артефактов

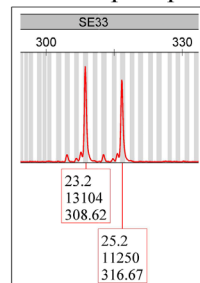
Побочные пики



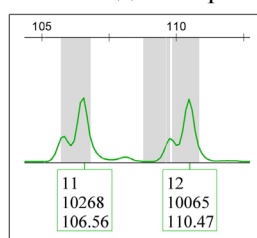
Гетерозиготный дисбаланс



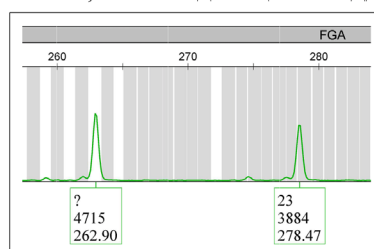
Аллель микроварианта



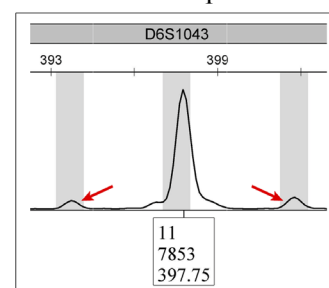
Неполное аденилирование



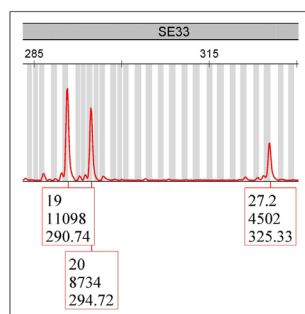
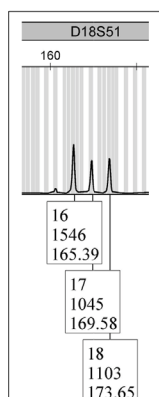
Аллель, не попадающая в лэддер



Статтеры



Триаллельные паттерны



Шум базовой линии

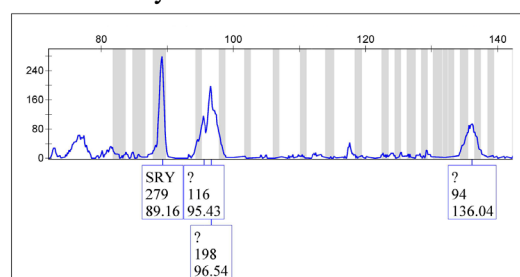


Рис. 1 Краткая схема распространенных артефактов анализа STR-CE, проиллюстрированных фрагментами электрофореграмм из лабораторной практики.

Примечание: STR-CE — short tandem repeat, products are separated by size using capillary electrophoresis (профилирование на основе коротких тандемных повторов с последующим анализом ампликонов методом капиллярного электрофоретического разделения). Цветное изображение доступно в электронной версии журнала.

за гетерогенности и анеуплоидии, встречающихся у многих иммортализованных и трансформированных клеток, электрофоретическая картина часто отличается от ожидаемых для соматических клеток человека 1-2 пиков на локус.

В настоящее время даже у одной и той же клеточной линии при изучении информации из разных источников, агрегированной международными базами данных (например, Cellosaurus), STR-профили в какой-то степени могут различаться. Причиной этого является как генетическая гетеро-

генность линии, культивируемой в условиях разных лабораторий на протяжении долгого времени, так и использование разных наборов для STR-профилирования [19].

С технической точки зрения работать с клеточными культурами проще, чем с криминалистическими образцами. Это объясняется тем, что в большинстве случаев не возникает проблем с недостатком или деградацией генетического материала и намного реже встречаются ситуации смесей ДНК в одной пробирке, т.е. применительно к исследова-

ниям на клеточных линиях — контаминация культуры клеток [13].

При соблюдении условий корректной постановки анализа и внимательном отношении к прибору и качеству реактивов, количество и интенсивность артефактов заметно снижены, зачастую в таком случае они вообще значимо не влияют и не затрудняют интерпретацию результатов анализа. Однако по разным причинам не всегда удается получить недвусмысленную электрофоретическую картину, поэтому умение работать с нестандартными результатами и отделять истинный сигнал от артефакта — важные навыки для оператора.

Самыми распространёнными артефактами STR-анализа являются шум, статтеры, побочные пики, просачивание красителя и гетерозиготный дисбаланс [20, 28]. Также встречаются капли краски, нематричное присоединение, аллели микровариантов, аллели, не попадающие в аллельную лестницу, выпадение аллелей (локусов) и триаллельные образцы [20].

Для каждого упомянутого артефакта ниже будет дано описание. Отдельные случаи возникновения артефактов, которые представляют наибольшее затруднение при интерпретации данных, будут рассмотрены подробно вместе с иллюстрациями из лабораторной практики и рекомендациями по решению проблем.

Стоит отметить, что наиболее часто встречающиеся и характерные артефакты при STR-профилеировании ДНК опухолевых клеточных линий и ДНК здорового человека (например, при установлении отцовства) различаются. В настоящей статье акцент сделан на артефакты при анализе опухолевых клеточных линий.

Шум. Шум представляет собой серию небольших невоспроизводимых пиков на уровне базовой линии, которые всегда присутствуют в результатах образцов в качестве фона и обычно не затрудняют анализ. Шум обусловлен электрическими флуктуациями в ходе электрофоретического разделения, наличием микропузырьков воздуха, мелких взвешенных частиц или кристаллов полимера в капиллярах, загрязнением в технических узлах генетического анализатора (камера насоса, окно детекции), состоянием капилляров и полимера, а также низкоуровневым загрязнением самих исследуемых образцов [29]. Мелкие взвешенные частицы, а также микропузырьки увеличивают амплитуду шума базовой линии, затрудняя распознавание пиков определяемых компонентов, поэтому до начала анализа электрофореграмм в программе для обработки данных автоматически или вручную определяется аналитический порог — LOD (Limit of Detection, предел обнаружения).

Аналитический порог (LOD) — это такое значение, выраженное в относительных единицах флу-

оресценции (Relative Fluorescence Unit, RFU), которое выражает статистический доверительный предел шумовой ошибки, обычно покрывающее 99,7% случаев (т.е. составляет три стандартных отклонения величины шума). Обычно LOD рассчитывается по формуле 1 [29]:

$$\text{LOD} = \mu_b + 3\sigma_b, (1)$$

где μ_b — среднее значение RFU сигнала базовой линии; σ_b — стандартное отклонение величины μ_b .

Или по формуле 2 [30]:

$$\text{ATin} = e^{v+kt}, (2)$$

где ATin — аналитический порог детекции, v — среднее значение RFU сигнала базовой линии, t — дисперсия натурального логарифма (логарифм по основанию e) шумовых сигналов, а k — числовой коэффициент, выбранный в соответствии с желаемым уровнем достоверности (обычно в диапазоне 3-5) [28].

Предел количественного определения — LOQ (Limit of Quantification), или стохастический предел, представляет собой величину, выраженную в относительных единицах флуоресценции — RFU (Relative Fluorescence Unit), которая задает минимальное значение интенсивности флуоресценции потенциального пика, который будет однозначно расцениваться за истинный сигнал [27]. Рассчитывается по формуле (3):

$$\text{LOQ} = \mu_b + 10\sigma_b, (3)$$

где μ_b — среднее значение RFU сигнала базовой линии; σ_b — стандартное отклонение величины μ_b .

Однако бывает, что значения RFU шумовых сигналов приближаются к заданному пороговому значению LOQ, в таком случае они могут быть приняты за пики и затруднить распознавание истинных сигналов. Самое важное отличие истинного пика от шумового заключается в том, что шумовой не воспроизводится. Поэтому, в случае получения сомнительных результатов, рекомендуется оценить электрофореграмму технического повтора исследуемого образца или провести повторный анализ.

Капли краски. На электрофореграмме такие пики имеют более пологую форму и широкое основание в отличие от истинных пиков. Обычно причиной таких капель, детектируемых на электрофореграмме, является разложение флуоресцентно-меченых праймеров вследствие их неправильного хранения или недостаточной очистки [20]. При корректной работе с коммерческими наборами реактивов данный артефакт встречается довольно редко, поскольку обычно производители проводят тщательную подготовку и очистку праймеров, входящих в состав набора, поэтому, при соблюдении условий хранения, целостность праймеров сохраняется на протяжении срока годности. На электрофореграмме пики от слипшихся фрагментов красителя можно достаточно точно отделить от истинных пиков: у таких пиков нетипичная широкая форма

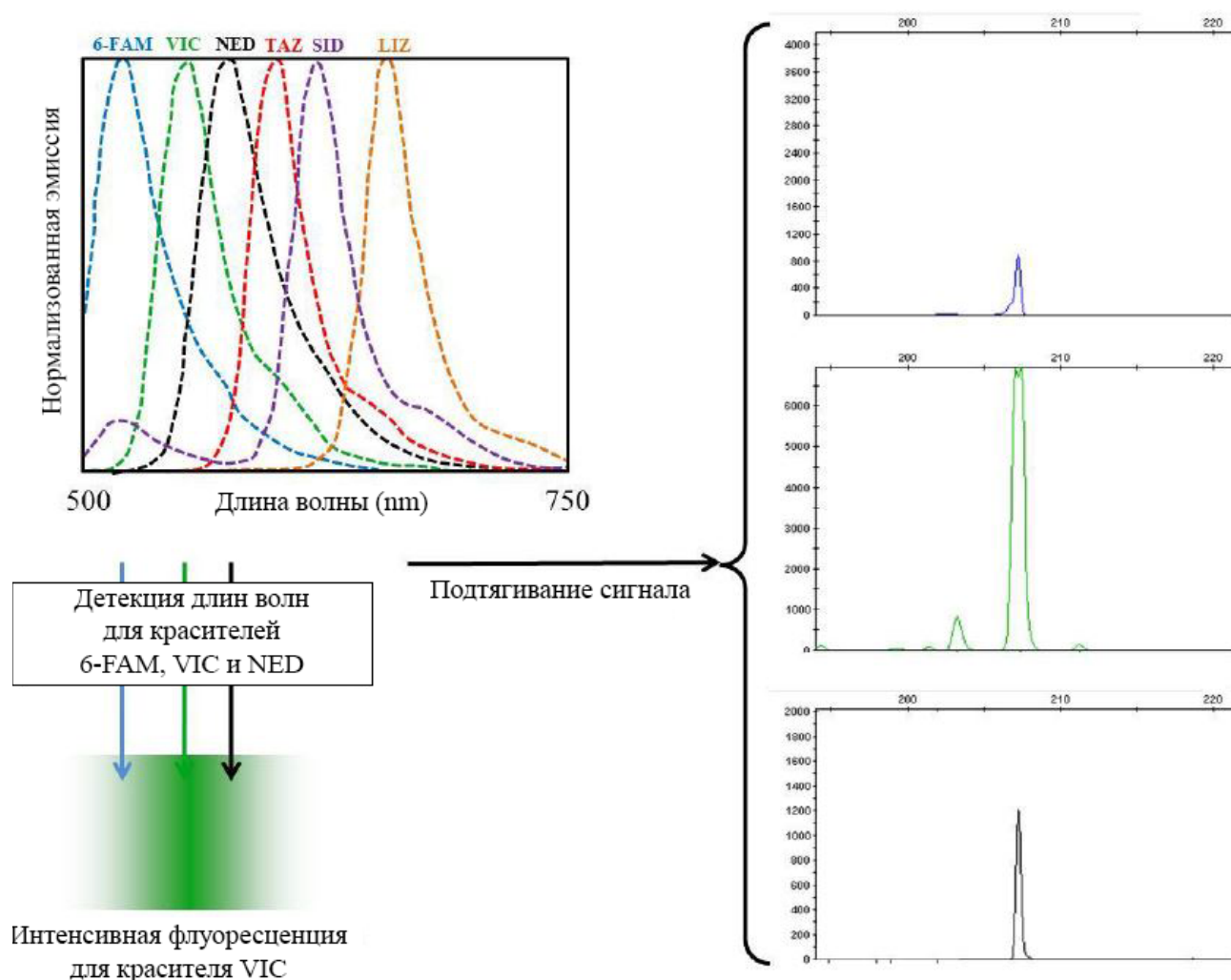


Рис. 2 Длины волн флуорофоров, использованных в наборе GlobalFiler PCR amplification kit (Life Technologies), демонстрируют пример подтягивания сигнала от красителя VIC (зеленый) в дорожки с красителями 6-FAM (голубой) и NED (черный). У электрофоретических графиков, показанных справа, вертикальная ось соответствует интенсивности сигнала в относительных единицах флуоресценции (RFU), а горизонтальная ось отображает размер ампликона количеством пар оснований (bp) (адаптировано из статьи авторов [32]).

Примечание: цветное изображение доступно в электронной версии журнала.

у основания, и они детектируются только в одном канале красителей.

Если же при работе оператор часто сталкивается с данным видом артефактов, то рекомендуется предварительная очистка амплифицированных смесей от остатков праймеров и побочных продуктов или разбавление наносимого на электрофорез ампликона в несколько раз для снижения количества таких капель краски.

Подтягивание/просачивание красителя. Как было упомянуто ранее, недостатком аналитических систем с многоканальной флуоресцентной детекцией является то, что у красителей, используемых в коммерческих наборах, диапазоны длин волн испускания флуоресценции в какой-то степени перекрываются. Поскольку прибор для капиллярного электрофореза имеет только один лазер для возбуждения флуоресценции, при детекции это при-

водит к просачиванию сигнала от одного флуорофора в канал соседнего. Для сглаживания данного явления ПО прибора математически нормализует выходные данные, разделяя полученный с ячейки детекции сигнал по отдельным каналам, вычитая долю перекрываемого сигнала [24, 31]. Это практически всегда позволяет корректно соотнести сигналы от разных флуорофоров и избежать проблем при анализе данных оператором, несмотря на большую потерю сигнала в некоторых каналах.

Просачивание красителя возникает, когда избыточное количество фрагментов, меченных определенным флуорофором, обнаруживается CCD-камерой, и на соответствующей этому красителю полосе электрофореграммы образуется пик крайне высокой интенсивности. В таком случае на дорожках соседних красителей со смежной длиной испускания флуоресценции в той же точке по горизон-

тальной оси будут наблюдаться побочные ложные пики нетипичной формы более низкой интенсивности [32].

Как это выглядит на электрофореграмме: крайне высокий пик в зеленом канале красителя и острый узкий пик в синем канале, регистрируемый в той же позиции по горизонтальной оси времени или длины нуклеотидной последовательности (рисунок 2).

Часто такой артефакт можно наблюдать, когда в смесь для ПЦР добавляется избыточное количество матричной ДНК [31]. Такие пики, если они имеют значительную высоту, могут остаться нераспознанными ПО как артефакты и быть принятыми за истинные аллели, поэтому важна дополнительная проверка результатов оператором. Важной особенностью идентификации данного вида артефактов является то, что небольшие пики-выбросы регистрируются только в соседних каналах детекции и строго в той же позиции по горизонтальной оси. Повторная амплификация с меньшим количеством вводимой ДНК или разведение в 10 раз исследованной пробы при внесении на планшет прибора может свести проявление данного артефакта к минимуму.

Статтеры. Статтеры (англ. Stutter) — это самые распространенные артефакты анализа [33]. Они образуются при амплификации как минорные побочные ПЦР-продукты, содержащие на кратное количество (один, реже два) больше или меньше микросателлитных повторов основного аллеля, и обозначаются как $n \pm 1, 2, 3$ -статтер, где n — истинное число микросателлитных повторов в образце.

Возможный механизм образования статтеров — некорректная транскрипция полимеразой целевого участка. Такая ситуация возникает, когда после начала синтеза нуклеотидной цепи с праймера матричная цепь и праймер по какой-то причине частично диссоциируют [34]. Если такое происходит во время ПЦР, но впоследствии праймер и матричная цепь повторно отжигаются с полной точностью и амплификация продолжается, то реакция дает только один целевой продукт. Неправильный повторный отжиг с образованием шпилек на матричной цепи или праймере и амплификация такой молекулы могут приводить к образованию и накоплению нескольких побочных продуктов длиной $n \pm 1, 2, 3 \dots$ наряду с целевым [20]. В этом случае на электрофореграмме рядом с истинными появляются небольшие дополнительные ложные пики, которые и называются статтерами.

Обычно, на частоту появления и высоту статтеров влияют такие факторы, как [20, 25]:

- количество нуклеотидов в единице повторов (динуклеотидные повторы характеризуются большим количеством статтеров, чем тетрануклеотидные);
- степень однородности и сложности тандемного повтора (у простых тандемных повторов высота статтеров выше, чем у сложных и составных);

- длина аллеля в локусе (чем больше повторов содержится в локусе, тем чаще детектируются статтеры аллелей).

Самый часто встречающийся статтер — это $n-1$, поскольку такой ампликон накапливается при однократном изменении длины матричной цепи или праймера во время частичной диссоциации, и это событие — самое энергетически выгодное из возможных для образования статтеров [20]. В основном (95% случаев), при амплификации достаточного количества ДНК высота статтеров не превышает порог в 15% [25, 35]. Соотношение высоты пика статтера к высоте истинного пика измеряется отношением высоты статтера к сумме высот этого статтера и истинного пика в % [36]. Чаще при интерпретации используют коэффициент статтера, который равен отношению высоты статтера к высоте истинного пика (в %) [34].

Программы для обработки данных STR-профилирования обычно имеют фильтр, который позволяет исключать пики статтеров, если отношение их высоты к высоте основного (истинного) пика $\leq 20\%$ [20, 35]. Некоторые программы позволяют выставить % порог фильтра вручную и индивидуально под каждый из маркеров, поскольку существует огромное количество исследований, направленных на изучение диапазона высот статтеров в зависимости от локуса, популяционных и этнических особенностей генотипа и др. [37].

При работе с клеточными линиями такая тщательная настройка отдельных для каждого локуса коэффициентов статтера не требуется, поскольку в сложных случаях почти всегда есть возможность повторить анализ, благодаря достаточному стартовому количеству ДНК и оптимальному качеству генетического материала, в отличие от криминалистических образцов.

Сложность при интерпретации результатов клеточных линий бывает в том случае, когда образец контаминирован или по какой-либо другой причине содержит ДНК из нескольких источников, тогда аллели минорного контрибьютора могут быть приняты за статтеры и наоборот; это определенно усложняет детекцию контаминации и определение истинного STR-профиля исследуемого образца.

Шипы, побочные пики. Это узкие, резкие и относительно высокие пики на графике, которые появляются из-за флуктуации напряжения в ходе анализа. Флуктуации напряжения могут быть вызваны присутствием пузырьков воздуха в капилляре, инородных частиц (например, пыли) или образованием кристаллов полимера или формамида. Обычно программы при обработке первичных данных хорошо справляются с задачей по распознаванию и исключению таких артефактов, но иногда их приходится исключать из анализа вручную.

Шипы, если они не появляются вдоль истинного пика, как правило, появляются в одной и той же точ-

ке данных во всех каналах красителей, эта характерная особенность позволяет отличить их от других артефактов. Появление таких всплесков можно свести к минимуму, используя свежий набор реагентов и образцов. Однако частое появление побочных пиков и шипов свидетельствует о неисправности системы.

Нематричное присоединение. Это известный артефакт ПЦР, который возникает из-за того, что некоторые полимеразы обладают внутренней, независимой от матрицы, полимеразной активностью, ответственной за добавление некодируемого матрицей остатка аденина, как правило, к вновь синтезированной цепи [20].

Полимераза Таq, например, после окончания синтеза дочерней цепи ДНК добавляет нуклеотиды уже без матрицы, в основном аденин (аденилирование), к 3'-концу продукта ПЦР, в результате чего некоторые продукты ПЦР становятся на одно основание больше, чем продукт от истинного аллеля. Причина такого поведения некоторых полимераз не ясна до конца, известно, что нематричное присоединение нуклеотидов, их количество и тип зависят от семейства полимеразы, от типа последнего нуклеотида в матрице, последовательности праймера, от расположения флуоресцентного красителя и от соотношения количества полимеразы и ДНК в реакционной смеси, а также от условий проведения ПЦР [38].

Аденилирование приводит к расщеплению пиков. Помимо соответствующего основному продукту, появляется дополнительный пик, расположенный на расстоянии ± 1 нуклеотид от основного. Тандем пиков в такой ситуации называют +А и -А.

Условия амплификации и полимеразы большинства коммерческих наборов реагентов для STR-профилирования хорошо сбалансированы за счет различных приёмов (особая нуклеотидная последовательность праймеров, добавление рестрикционных ферментов, модифицированная полимеразы), поэтому такой артефакт, как нематричное присоединение, встречается довольно редко в анализе, и ампликоны имеют одинаковую длину. В таких случаях причиной неполного аденилирования может стать высокая начальная концентрация ДНК относительно реакционной ёмкости ПЦР-смеси.

Если при анализе продуктов мультиплексной ПЦР на электрофореze все-таки обнаруживается картина +А и -А пиков, то чтобы предотвратить такое дробление сигнала и способствовать полному аденилированию ампликонов, обычно к заключительному этапу ПЦР добавляют 10–60 мин инкубации образцов при 60 или 72°С [20]. Если причиной неполного аденилирования стала высокая концентрация внесенной в реакцию ДНК, то такие образцы можно разбавить буфером для ПЦР, добавить полимеразы и провести элонгацию в ПЦР-амплификаторе.

Аллели, не попадающие в диапазон аллельной лестницы. Это пики, имеющие форму истинных

пиков аллелей и находящиеся в диапазоне нуклеотидного размера конкретного маркера, которые выбраковываются или не определяются ПО для анализа электрофореграмм из-за непредставленности в аллельной лестнице (reference ladder) [39]. Входящая в любой коммерческий набор по STR-профилированию аллельная лестница содержит в себе смесь аллелей, которые наиболее широко встречаются по популяции [20]. По частоте встречаемости тех или иных аллелей в различных популяциях написано много статей, на эти данные опираются и при разработке наборов реагентов [20, 25, 32, 40–43].

Однако человеческая популяция может содержать редкие аллели, которые отличны от входящих в состав набора реактивов. При анализе клеточных линий, особенно опухолевых, также встречаются нетипичные аллели, что связано с большей скоростью их генетических изменений или межвидовой контаминацией образца.

Редкий аллель, который все-таки амплифицируется, а соответствующий ему пик не попадает в границы диапазона заданных коммерческой панелью аллелей, называется аллелем, не попадающим в диапазон аллельной лестницы (off-ladder allele, OL) [20]. В таких случаях полученный результат подтверждают повторной постановкой образца на анализ. Если результат воспроизводится, а в литературе или базах данных есть информация о существовании такого аллеля, его можно считать истинным. В некоторых ПО можно задать значение такого аллеля (т.е. отобразить численно количество тандемных повторов), но это приходится делать вручную при анализе результатов оператором.

Если при повторном анализе пик не воспроизводится, следует заподозрить другой вид артефакта, например, побочный пик, шип, каплю краски, шум или внутривидовую контаминацию.

Аллели микровариантов. Иногда при анализе электрофореграмм можно столкнуться с истинными пиками, которые не представлены в лэддере, а количество тандемных повторов таких детектируемых фрагментов представляет собой не кратное число повторяющихся элементов, а, например, 14 целых единиц повтора (AGAA) и один неполный повтор (AG) в каком-либо месте последовательности. В таком случае возникает микровариантная аллель 14.2, что соответствует 14 полным повторам и 2 нуклеотидам неполного. Чаще всего такие микроварианты встречаются в высокополиморфных локусах, самые распространенные микроварианты включены в состав аллельной лестницы и не представляют проблем при анализе [20, 25].

При изучении оператором электрофореграмм такие нетипичные (неопределяемые ПО) микроварианты аллелей можно задать вручную после исключения других возможных причин для появления рассматриваемого артефакта и повторного

анализа. Надежным доказательством для подтверждения данного микроварианта служит секвенирование целевого участка. При анализе опухолевых клеточных линий появление микровариантов не редко, поскольку клетки предрасположены к хромосомным перестройкам, поломкам и менее драматичным изменениям генома.

Выпадение аллеля (выпадение аллельных вариантов). Это случай, когда в локусе не наблюдается сигнала от одного из аллелей (если таковой предполагается, например, если профиль клеточной культуры известен по генетическому паспорту заранее) или вообще отсутствуют пики накопления продукта ПЦР. Причин этому может быть много, их можно разделить на две категории: генетические изменения и технические проблемы анализа.

К основным генетическим причинам можно отнести мутации в сайтах связывания праймера, а для опухолевых клеточных линий следует упомянуть потерю и поломку частей или целых хромосом, что также может приводить к потере аллелей [44, 45]. Технические причины выпадения аллеля могут быть связаны с проблемами амплификации из-за деградированной ДНК (распространенная проблема у криминалистов) или ингибиторов ПЦР. Ингибиторы ПЦР могут попасть в реакционную смесь вместе с биоматериалом или в процессе выделения ДНК и не до конца элиминироваться в процессе очистки препарата нуклеиновой кислоты.

Поскольку современные методы выделения ДНК из клеточных культур в лаборатории позволяют выделять препараты ДНК, удовлетворяющие требованиям STR-анализа, в случае выпадения аллелей при анализе клеточных культур влияние генетического фактора более вероятно. Это можно подтвердить повторной постановкой образца, а чтобы исследовать причину такого результата, необходимо применение дополнительных методов (секвенирование фрагмента последовательности, кариотипирование и др.).

Триаллельный образец. В некоторых случаях можно наблюдать три аллеля в локусе STR-профиля аутосомной ДНК в образце из одного источника [45, 46]. При анализе ДНК человека данный артефакт может быть обусловлен мутацией соматических клеток на раннем этапе развития эмбриона или событиями хромосомной дупликации и перестройки в зародышевой линии, но также может наблюдаться, например, при синдроме Дауна или других генетических особенностях индивида [46, 47]. В случае анализа STR-профилей опухолевых образцов, полученные данные следует интерпретировать исходя из особенностей возможной генетической организации опухолевых клеток: наличие нескольких ядер, наличие лишнего фрагмента хромосомы или целой хромосомы, многочисленные мутации или сосуществование нескольких субклонов в клеточном флаконе.

Было установлено два типа триаллельных паттернов [20]:

- 1 тип, где сумма высот пиков двух аллелей равна высоте пика третьего аллеля (встречается чаще по сравнению с 2 типом [48]);
- 2 тип, где высота пиков трех аллелей уравновешена по своей высоте.

При анализе электрофореграмм клеточных линий в лаборатории наличие "лишних" (когда присутствует >2) пиков в нескольких локусах может свидетельствовать о смешанном образце ДНК из разных источников, например, из-за наличия нескольких субклонов или внутривидовой контаминации. Однако, если картина трёх пиков регистрируется только в одном локусе, вероятнее всего, это генетически обусловленный артефакт.

В некоторых случаях может наблюдаться ложный триаллельный паттерн, когда крайний пик вне аллельной лестницы от соседнего маркера неправильно интерпретирован. Кроме того, иногда триаллельный паттерн 1 типа может представлять собой 2 истинных аллеля и ложно принятый за аллель крайне высокий статтер между ними, полученный суммарным побочным продуктом обоих аллелей, если они расположены рядом, например, 14, 15 и 16 аллели. На такие случаи следует обращать внимание при интерпретации данных.

Гетерозиготный дисбаланс. Поскольку диплоидный набор хромосом человека состоит из пар соответствующих хромосом, можно было бы предположить, что амплификация ПЦР-продуктов с локуса на каждой хромосоме из пары должна проходить равноценно, соответственно, количество ПЦР-продукта и величина детектируемого сигнала от каждой аллели при STR-профилировании должны быть примерно равны. На деле же, взаимосвязь между количеством исходной матрицы и высотой пиков аллелей имеет стохастический характер. Это объясняется следующими факторами: эффективностью экстракции и амплификации, стохастическим распределением молекул ДНК при отборе пробы дозатором на всех этапах пробоподготовки, а также длиной и нуклеотидной последовательностью амплифицируемых фрагментов [25].

Следовательно, мы наблюдаем картину гетерозиготного дисбаланса сигналов, которая выглядит так: на электрофореграмме в одном локусе два пика с довольно большой разницей высот (например, высота аллеля 2 составляет всего 68% от высоты аллеля 1).

Гетерозиготный дисбаланс наблюдается всегда, но в разной степени выраженности. Был подсчитан допустимый ("безопасный" для принятия решений о статусе сигнала) диапазон гетерозиготного дисбаланса, и он составляет 0,6-1,66. Рассчитать данный параметр можно, применяя наиболее распространенную формулу для подсчета гетерозиготного дисбаланса (4), эта рекомендация выглядит так:

$$0,6 \leq H_b = \frac{\phi_{HMM}}{\phi_{LMW}} \leq 1,66, (4)$$

где H_b — степень гетерозиготного дисбаланса; ϕ_{HMM} — величина сигнала пика аллеля локуса с большим молекулярным весом (RFU); ϕ_{LMW} — величина сигнала низкого пика локуса с меньшим молекулярным весом (RFU).

Таким образом, при наблюдении гетерозиготного дисбаланса, попадающего в заданный диапазон, мы можем валидировать обе наблюдаемые аллели. Однако если разница высот пиков аллелей значительна, стоит подвергнуть сомнению такое простое разрешение электрофоретической картины, потому что причиной может быть целый спектр событий: неэффективная амплификация аллеля из-за мутаций в праймерной области, контаминация посторонней ДНК, побочный пик и т.д.

Заключение

Артефакты анализа — это, несомненно, проблема и неотъемлемая часть работы с данным молекулярно-генетическим методом. Их влияние на конечный результат можно свести к минимуму при использовании реактивов высокого качества, корректных концентраций ДНК в оптимальных диапазонах, заданных производителями наборов для исследования, при постоянном техническом обслуживании прибора.

Однако эти условия не всегда удается поддерживать в лаборатории, поэтому встреча с артефактами неизбежна, и зачастую это означает, что оператор должен вручную проверять полученные результаты. Тем не менее, зная характерные особенности каж-

дой группы артефактов, можно довольно эффективно интерпретировать полученные данные и получать надежные результаты.

Кроме того, стоит обратить внимание на положительный аспект данного явления — артефакты могут рассматриваться операторами как дополнительный качественный индикатор состояния всей системы, поэтому, зная причины их возникновения и характерные особенности, можно не только безошибочно детектировать и исключать их, получая корректные результаты, но и определять уязвимости в процессе работы.

В рамках проекта по созданию биобанка на базе НМИЦ онкологии им. Н. Н. Блохина нами проведена значительная работа по STR-профилированию образцов клеточных культур. По ходу работы мы сталкивались с проявлениями основных, затрудняющих интерпретацию результатов, артефактов анализа. Были опробованы найденные в литературе методики варьирования условий анализа, которые позволили сгладить негативный эффект артефактов на результаты. Таким образом, можно заключить, что наши эмпирические наблюдения согласуются с данными, приведенными в литературе.

Отношения и деятельность. Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках научно-исследовательской работы № 075-15-2021-1060 "Создание и развитие биоресурсной коллекции генетически и фенотипически охарактеризованных клеточных линий и первичных опухолей человека".

Литература/References

- Nardone RM. Eradication of cross-contaminated cell lines: a call for action. *Cell Biol Toxicol.* 2007;23(6):367-72. doi:10.1007/s10565-007-9019-9.
- Freedman LP, Gibson MC, Ethier SP, et al. Reproducibility: changing the policies and culture of cell line authentication. *Nat Methods.* 2015;12(6):493-7. doi:10.1038/nmeth.3403.
- Freedman LP, Gibson MC, Wisman R, et al. The culture of cell culture practices and authentication — Results from a 2015 Survey. *Biotechniques.* 2015;59(4):189-90. doi:10.2144/000114344.
- Bian X, Yang Z, Feng H, et al. A Combination of Species Identification and STR Profiling Identifies Cross-contaminated Cells from 482 Human Tumor Cell Lines. *Sci Rep.* 2017;7:9774. doi:10.1038/s41598-017-09660-w.
- Souren NY, Fusenig NE, Heck S, et al. Cell line authentication: a necessity for reproducible biomedical research. *EMBO J.* 2022; 41(14):e111307. doi:10.15252/embj.2022111307.
- Lin LC, Elkashty O, Ramamoorthi M, et al. Cross-contamination of the human salivary gland HSG cell line with HeLa cells: A STR analysis study. *Oral Dis.* 2018;24(8):1477-83. doi:10.1111/odi.12920.
- Horbach SPJM, Halfman W. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLoS One.* 2017;12(10):e0186281. doi:10.1371/journal.pone.0186281.
- Ye F, Chen C, Qin J, et al. Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. *FASEB J.* 2015;29(10):4268-72. doi:10.1096/fj.14-266718.
- Korch C, Hall EM, Dirks WG, et al. Authentication of M14 melanoma cell line proves misidentification of MDA-MB-435 breast cancer cell line. *Int J Cancer.* 2018;142(3):561-72. doi:10.1002/ijc.31067.
- Almeida JL, Cole KD, Plant AL. Standards for Cell Line Authentication and Beyond. *PLoS Biol.* 2016;14(6):e1002476. doi:10.1371/journal.pbio.1002476.
- Vicente E, Lesniewski M, Newman D, et al. Best Practices for Authentication of Cell Lines to Ensure Data Reproducibility and Integrity. *Radiat Res.* 2022;197(3):209-17. doi:10.1667/RADE-21-00148.1.
- Yu M, Selvaraj SK, Liang-Chu MM, et al. A resource for cell line authentication, annotation and quality control. *Nature.* 2015; 520(7547):307-11. doi:10.1038/nature14397.
- Kosobokova EN, Malchenkova AA, Kalinina NA, et al. Using short tandem repeat profiling to validate cell lines in biobanks. *Cardiovascular Therapy and Prevention.* 2022;21(11):3386. (In Russ.) Кособокова Е. Н., Мальченкова А. А., Калинина Н. А. и др. Использование метода профилирования на основе коротких tandemных повторов для подтверждения подлинности клеточных линий в биобанках. *Кардиоваскулярная терапия и профилактика.* 2022;21(11):3386. doi:10.15829/1728-8800-2022-3386.
- Anisimov SV, Axmerov TM, Balanovskij OP, et al. Biobanking. National Guide. Moscow: ООО "Izdatel'stvo TRIUMF", 2022. 308 p. (In Russ.) Анисимов С. В., Ахмеров Т. М., Балановский О. П. и др. Биобанкирование. Национальное руководство. М.: ООО "Издательство ТРИУМФ", 2022. 308 с. ISBN 978-5-93673-322-2.

15. Parson W, Kirchbner R, Mühlmann R, et al. Cancer cell line identification by short tandem repeat profiling: power and limitations. *FASEB J*. 2005;19(3):434-6. doi:10.1096/fj.04-3062fje.
16. Almeida JL, Hill CR, Cole KD. Mouse cell line authentication. *Cytotechnology*. 2014;66(1):133-47. doi:10.1007/s10616-013-9545-7.
17. Almeida JL, Dakic A, Kindig K, et al. Interlaboratory study to validate a STR profiling method for intraspecies identification of mouse cell lines. *PLoS One*. 2019;14(6):e0218412. doi:10.1371/journal.pone.0218412.
18. Almeida JL, Hill CR, Cole KD. Authentication of African green monkey cell lines using human short tandem repeat markers. *BMC Biotechnol*. 2011;11:102. doi:10.1186/1472-6750-11-102. PMID: 22059503; PMCID: PMC3221628.
19. Chen YH, Connelly JP, Florian C, et al. Short tandem repeat profiling via next-generation sequencing for cell line authentication. *Dis Model Mech*. 2023;16(10):dmm050150. doi:10.1242/dmm.050150.
20. Butler JM. "Chapter 3 — STR alleles and amplification artifacts," in *Advanced topics in forensic DNA typing: Interpretation*. Editor JM Butler (San Diego: Academic Press, 2015:47-86). ISBN: 978-0-12-405213-0.
21. Gilbert DA, Reid YA, Gail MH, et al. Application of DNA fingerprints for cell-line individualization. *Am J Hum Genet*. 1990;47(3):499-514.
22. Bukreev YuM, Kosobokova EN, Kardashova SS, et al. DNA-markers for genetic monitoring of laboratory mouse. *Russian Journal of Biotherapy*. 2017;16(3):86-91. (In Russ.) Букреев Ю.М., Кособокова Е.Н., Кардашова С.С. Мониторинг чистоты линий лабораторных мышей с использованием ДНК-маркеров. *Российский биотерапевтический журнал*. 2017;16(3):86-91. doi:10.17650/1726-9784-2017-16-3-86-91.
23. Goor RM, Forman Neall L, Hoffman D, et al. A mathematical approach to the analysis of multiplex DNA profiles. *Bull Math Biol*. 2011;73(8):1909-31. doi:10.1007/s11538-010-9598-0.
24. Mohammed O, Assaleh K, Hussein G, et al. Novel algorithms for accurate DNA base-calling. *J Biomed Sci Eng*. 2013;6:165-74. doi:10.4236/jbise.2013.62020.
25. Gill P, Haned H, Bleka O, et al. Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches — Twenty years of research and development. *Forensic Sci Int Genet*. 2015;18:100-17. doi:10.1016/j.fsigen.2015.03.014.
26. Puch-Solis R, Pope S. Interpretation of DNA data within the context of UK forensic science — evaluation. *Emerg Top Life Sci*. 2021;5(3):405-13. doi:10.1042/ETLS20200340.
27. Gill P, Gusmão L, Haned H, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Sci Int Genet*. 2012;6(6):679-88. doi:10.1016/j.fsigen.2012.06.002.
28. Karkar S, Alfonse LE, Grgicak CM, et al. Statistical modeling of STR capillary electrophoresis signal. *BMC Bioinformatics*. 2019;20:584. doi:10.1186/s12859-019-3074-0.
29. Gilder JR, Doom TE, Inman K, et al. Run-specific limits of detection and quantitation for STR-based DNA testing. *J Forensic Sci*. 2007;52(1):97-101. doi:10.1111/j.1556-4029.2006.00318.x.
30. Mönich UJ, Duffy K, Médard M, et al. Probabilistic characterisation of baseline noise in STR profiles. *Forensic Sci Int Genet*. 2015;19:107-22. doi:10.1016/j.fsigen.2015.07.001.
31. Goor RM, Hoffman D, Riley GR. Novel Method for Accurately Assessing Pull-up Artifacts in STR Analysis. *Forensic Sci Int Genet*. 2021;51:102410. doi:10.1016/j.fsigen.2020.102410.
32. Taylor D, Powers D. Teaching artificial intelligence to read electropherograms. *Forensic Sci Int Genet*. 2016;25:10-8. doi:10.1016/j.fsigen.2016.07.013.
33. Vilsen SB, Tvedebrink T, Eriksen PS, et al. Stutter analysis of complex STR MPS data. *Forensic Sci Int Genet*. 2018;35:107-12. doi:10.1016/j.fsigen.2018.04.003.
34. Agudo MM, Aanes H, Roseth A, et al. A comprehensive characterization of MPS-STR stutter artefacts. *Forensic Sci Int Genet*. 2022;60:102728. doi:10.1016/j.fsigen.2022.102728.
35. Duffy KR, Gurram N, Peters KC, et al. Exploring STR signal in the single- and multicopy number regimes: Deductions from an in silico model of the entire DNA laboratory process. *Electrophoresis*. 2017;38(6):855-68. doi:10.1002/elps.201600385.
36. Taylor D, Bright JA, McGoven C, et al. Validating multiplexes for use in conjunction with modern interpretation strategies. *Forensic Sci Int Genet*. 2016;20:6-19. doi:10.1016/j.fsigen.2015.09.011.
37. Purps J, Siegert S, Willuweit S, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet*. 2014;12(100):12-23. doi:10.1016/j.fsigen.2014.04.008.
38. Zhou Y, Bo F, Tian T, et al. Excessive addition split peak formed by the non-templated nucleotide addition property of Taq DNA polymerase after PCR amplification. *Front Bioeng Biotechnol*. 2023;11:1180542. doi:10.3389/fbioe.2023.1180542.
39. Gorden EM, Sturk-Andreaggi K, Warnke-Sommer J, et al. Next generation sequencing of STR artifacts produced from historical bone samples. *Forensic Sci Int Genet*. 2020;49:102397. doi:10.1016/j.fsigen.2020.102397.
40. Song B, Fu J, Guo K. A Tibetan group from Ngawa Tibetan and Qiang Autonomous Prefecture, southwest China, is rich in genetic polymorphisms at 36 autosomal STR loci and shares a complex genetic structure with other Chinese populations. *Heliyon*. 2023;9(12):e23005. doi:10.1016/j.heliyon.2023.e23005.
41. Bottinelli M, Gouy A, Utz S, et al. Population genetic analysis of 12 X-chromosomal STRs in a Swiss sample. *Int J Legal Med*. 2022;136(2):561-3. doi:10.1007/s00414-021-02684-y.
42. Ashirbekov Y, Seidualy M, Abaildayev A, et al. Genetic polymorphism of Y-chromosome in Kazakh populations from Southern Kazakhstan. *BMC Genomics*. 2023;24(1):649. doi:10.1186/s12864-023-09753-z.
43. Lang Y, Guo F, Niu Q. StatsX v2.0: the interactive graphical software for population statistics on X-STR. *Int J Legal Med*. 2019;133(1):39-44. doi:10.1007/s00414-018-1824-6.
44. Risinskaya NV, Kozhevnikova Ya, Kovaleva V, et al. Loss of heterozygosity in the short tandem repeat (STR) profile of tumor DNA in patients with de novo diagnosed acute lymphoblastic leukemia as a pattern of abnormal tumor karyotype. *Cellular Therapy and Transplantation*. 2020;9(3):22-3. (In Russ.) Рисинская Н.В., Кожевникова Я., Ковалева В. и др. Потеря гетерозиготности в профиле коротких tandemных повторов (STR) опухолевой ДНК у пациентов с de novo диагностированным острым лимфобластным лейкозом как паттерн аномального кариотипа опухоли. *Клеточная терапия и трансплантация*. 2020;9(3):22-3. doi:10.18620/ctt-1866-8836-2020-9-3-1-152. EDN VSSEKO.
45. Kosobokova EN, Kalinina NA, Konoplina KM, et al. Human Metastatic Melanoma Cell Lines Panel for In Vitro and In Vivo Investigations. *J Mol Pathol*. 2024;5(1):11-27. doi:10.3390/jmp5010002.
46. Fan J, Zhang AP, Zheng ZZ, et al. A Case of Type 1 Triallelic Patterns at D5S818, D18S51, D6S1043, and FGA Demonstrated by Short Tandem Repeat Analysis. *Int J Clin Pract*. 2022;2022:8600125. doi:10.1155/2022/8600125.
47. Gymrek M. A genomic view of short tandem repeats. *Curr Opin Genet Dev*. 2017;44:9-16. doi:10.1016/j.gde.2017.01.012.
48. Yang Q, Shen Y, Shao C, et al. Genetic analysis of tri-allelic patterns at the CODIS STR loci. *Mol Genet Genomics*. 2020;295(5):1263-8. doi:10.1007/s00438-020-01701-w.