Лекция 8. Методы секвенирования генома, автоматизация данного процесса.

1 Методы секвенирования генома.

2 Автоматизация процесса секвенирования.

1. Методы секвенирования генома. Появление секвенирования нового поколения (NGS) привело к смене парадигмы в геномных исследованиях, предлагая непревзойденные возможности для анализа молекул ДНК и РНК высокопроизводительным и экономически эффективным способом. Эта преобразующая технология быстро продвинула достижения геномики в различных областях. NGS позволяет быстро секвенировать миллионы фрагментов ДНК одновременно, предоставляя всестороннее понимание структуры генома, генетических вариаций, профилей экспрессии генов и эпигенетических модификаций. Универсальность платформ NGS расширила сферу геномных исследований, способствуя исследованиям редких генетических заболеваний, геномике рака, анализу микробиома, инфекционным заболеваниям и популяционной генетике. Более того, NGS позволил разработать таргетную терапию, подходы точной медицины и улучшенные методы диагностики. В этом обзоре представлен глубокий обзор текущих тенденций и последних достижений в технологии NGS, подчеркивая ее потенциальное влияние на различные области геномных исследований. Кроме того, в обзоре рассматриваются возникшие проблемы и будущие направления развития технологии NGS, включая усилия по повышению точности и чувствительности данных секвенирования, разработку новых алгоритмов анализа данных и поиск более эффективных, масштабируемых и экономически выгодных решений, которые ждут нас в будущем.

Секвенирование следующего поколения (NGS) произвело революцию в геномике, расширив наши знания о структуре, функции и динамике генома. Эта новаторская технология позволила проводить обширные исследования и позволила ученым изучать сложности генетической информации беспрецедентными способами. Благодаря своей высокой пропускной способности и экономической эффективности NGS стал основополагающим инструментом для исследователей в различных дисциплинах, от базовой биологии до клинической диагностики. NGS не только обеспечило комплексное секвенирование генома, но и облегчило транскриптомику, эпигеномику, метагеномику и другие исследования в области омики. Появление передовых платформ NGS, таких как Illumina, Pacific Biosciences и Oxford Nanopore, преобразило область геномики, позволив проводить параллельное секвенирование миллионов и миллиардов фрагментов ДНК. Эта возможность открыла новые возможности для понимания генетических вариаций, экспрессии генов, эпигенетических модификаций и микробного разнообразия. NGS сыграл важную роль в выявлении вариантов, вызывающих заболевания, открытии новых лекарственных мишеней и проливании света на сложные биологические явления, включая гетерогенность опухолей и процессов развития.

#### *Технология секвенирования первого поколения.* Первые попытки секвенирования ДНК и РНК включали химическую деградацию или ферментативное расщепление молекул для получения фрагментов, которые можно было бы анализировать по отдельности. Роберт Холли был первым, кто секвенировал молекулу нуклеиновой кислоты, тРНК аланина, в 1964 году с использованием рибонуклеазы из *S. cerevisiae*. Аналогичным образом Уолтер Гилберт и Аллан Максам разработали метод химической деградации, который позволил секвенировать полный бактериофаг PhiX174. Однако настоящий прорыв произошел с введением метода секвенирования на основе терминации цепи Фредериком Сэнгером. Этот метод использовал дидезоксинуклеотиды, которые прекращают удлинение цепи ДНК во время репликации, и позволял получать считывания последовательностей длиной до нескольких сотен нуклеотидов. Метод Сэнгера получил широкое распространение и произвел революцию в области молекулярной биологии, позволив быстро секвенировать ДНК и РНК. В 1987 году в США был запущен первый коммерческий автоматизированный секвенатор Applied Biosystems ABI 370. Этот аппарат использовал флуоресцентно меченые дидезоксинуклеотиды и капиллярный электрофорез для автоматизации метода секвенирования по Сэнгеру, значительно увеличив скорость и точность секвенирования ДНК. ABI 370 быстро стал отраслевым стандартом, а последующие усовершенствования технологии привели к разработке секвенаторов с более высокой пропускной способностью, способных производить более длинные считывания. Хотя технология первого поколения была в значительной степени вытеснена более новыми, более производительными технологиями секвенирования, она остается важной исторической вехой в развитии методов секвенирования. Возможность секвенировать ДНК и РНК произвела революцию во многих областях биологии и медицины и привела к многочисленным открытиям и достижениям в понимании генетики и молекулярной биологии.

#### *Технологии секвенирования второго поколения.* Методы секвенирования второго поколения произвели революцию в секвенировании ДНК, позволив одновременно секвенировать тысячи и миллионы фрагментов ДНК. Эти методы отличаются от традиционного секвенирования по Сэнгеру своей способностью выполнять параллельное секвенирование. Появилось несколько широко используемых платформ секвенирования второго поколения, одной из которых является метод секвенирования 454 компании Roche, который основан на пиросеквенировании, где последовательность определяется путем обнаружения высвобождения пирофосфата при добавлении нуклеотидов к шаблону ДНК. Другая платформа — секвенирование Ion Torrent, которое обнаруживает высвобождение ионов водорода во время синтеза ДНК для определения последовательности. Широко используемая платформа секвенирования Illumina использует метод секвенирования путем синтеза на основе обратимых терминаторов красителей. Другая перспективная технология, секвенирование SOLiD (секвенирование с помощью лигирования и обнаружения олигонуклеотидов), использует подход на основе лигирования с использованием обратимых терминаторов для определения последовательности ДНК. Эти технологии секвенирования второго поколения значительно увеличили пропускную способность и скорость секвенирования ДНК, что позволило использовать широкий спектр приложений в геномных исследованиях и клинической диагностике. Эти платформы сделали возможным секвенирование всего генома, анализ транскриптома и целевое секвенирование, что привело к прорывам в области генетической изменчивости, исследованиях заболеваний и персонализированной медицине. За эти годы было достигнуто множество достижений в методах секвенирования второго поколения.

#### Биология 12 00997 g002

#### *Рисунок 1. Обзор различных технологий NGS с разными платформами и принципами.*

#### *Секвенирование третьего поколения.* Технологии секвенирования третьего поколения представляют собой новейшие достижения в области секвенирования ДНК, предлагая новые подходы, которые преодолевают ограничения предыдущих поколений. Эти технологии обеспечивают возможности длинных прочтений секвенирования, позволяя секвенировать гораздо большие фрагменты ДНК по сравнению с более ранними методами. Примерами являются PacBio Sequencing, которое использует подход с одной молекулой в реальном времени (SMRT) с флуоресцентно мечеными нуклеотидами, что позволяет проводить длинную секвенцию фрагментов ДНК длиной до десятков килооснований. Другая технология — секвенирование Oxford Nanopore, основанное на технологии нанопор, где одноцепочечная молекула ДНК проходит через нанопору, и для определения последовательности ДНК измеряются изменения электрического тока. Секвенирование Oxford Nanopore обеспечивает длинные прочтения, портативность и анализ в реальном времени.

####

#### *Секвенирование длинных и коротких прочтений.* Основной принцип секвенирования с короткими прочтениями включает секвенирование путем синтеза на основе обогащения посредством гибридизации, амплификации или фрагментации. В то время как секвенирование с длинными прочтениями работает над обнаружением последовательности либо путем синтеза, либо путем изменения электрического напряжения/импеданса, генерируя ток, когда одно основание проходит через пору биологической мембраны. Секвенирование с длинными прочтениями может генерировать прочтения до 25–30 кб, тогда как секвенирование с короткими прочтениями может генерировать прочтения около 600–700 п.н. Кроме того, смещение амплификации устраняется в секвенировании с длинными прочтениями в отличие от секвенирования с короткими прочтениями. Поскольку подготовка библиотеки не требует ПЦР, модификация оснований, такая как метилирование ДНК, может быть легко обнаружена секвенированием с длинными прочтениями. Внедрение платформ высокопроизводительного секвенирования значительно снизило частоту ошибок и заметно повысило точность технологий секвенирования с длинными прочтениями. Секвенирование с коротким прочтением полезно для определения распространенности специфических последовательностей, профилирования экспрессии транскриптов и идентификации вариантов. Однако технологии секвенирования с длинным прочтением превосходят другие в обеспечении всестороннего покрытия генома, позволяя исследователям идентифицировать сложные структурные варианты, такие как большие вставки, делеции, инверсии, дупликации и многое другое.

## **Омики нового поколения на основе секвенирования.** Понимание сложных заболеваний человека требует интеграции данных из множества омических методов, таких как геномика, транскриптомика, эпигеномика и протеомика. Здесь мы кратко описываем различные омические технологии, которые реализованы на платформе NGS:

Геномные исследования с использованием NGS позволяют глубоко анализировать ДНК, используя различные подходы, такие как секвенирование всего генома, секвенирование всего экзома и целевое секвенирование.

#### *Полногеномное секвенирование.* Секвенирование всего генома (WGS) — это мощный и всеобъемлющий метод геномного анализа, который включает определение полной последовательности ДНК генома человека. Он предоставляет подробную схему генетического состава человека, охватывающую все гены, регуляторные области и некодирующие элементы, присутствующие в его геноме. Он находит свое применение в основном в научных открытиях, таких как исследования растений и животных, исследования рака, редких генетических заболеваний, пациентов со сложными симптомами заболеваний, популяционной генетике и новой сборке генома эукариот и прокариот. Секвенируя всю ДНК в геноме организма, WGS позволяет идентифицировать генетические вариации, начиная от однонуклеотидных полиморфизмов (SNP) до более крупных структурных изменений, таких как вставки, делеции и перестройки. Это богатство информации, полученной с помощью WGS, предлагает множество применений в различных областях. WGS имеет два типа подходов к секвенированию на основе размера генома, а именно: (1) крупное секвенирование всего генома, расшифровывающее более крупные геномы размером >5 Мб, такие как эукариоты, и (2) малое секвенирование всего генома, расшифровывающее меньшие геномы размером <5 Мб, в основном прокариоты. Короткое секвенирование предпочтительнее для вызова мутаций, в то время как длинное секвенирование предпочтительнее для сборки генома. Объединение короткого и длинного секвенирования для секвенирования новых геномов успешно применялось для точной сборки генома без референсной последовательности.

#### *Полноэкзомное секвенирование*. Секвенирование всего экзома (WES) — это подход к секвенированию, который фокусируется на захвате и секвенировании областей генома, кодирующих белки, известных как экзом. Экзом составляет приблизительно 1–2% всего генома, но содержит большинство известных вариантов, вызывающих заболевания. Секвенируя экзом, WES позволяет идентифицировать генетические вариации, включая однонуклеотидные варианты (SNV), вставки, делеции и вариации числа копий (CNV) в генах, кодирующих белки. WES является экономически эффективной альтернативой WGS для редких клинических заболеваний с кластерами симптомов, а также для идентификации вариантов для популяционной и раковой генетики. WES включает обогащение экзонных областей с использованием методов гибридного захвата или специфической для мишени амплификации с последующим высокопроизводительным секвенированием. Доступны различные анализы захвата экзома от NimbleGen, Agilent, Illumina, Twist и IDT, совместимые с платформой Illumina NGS. Биоинформационный подход, используемый для анализа данных WES, такой же, как и в WGS, поскольку WES является частью WGS.

#### *Целевое секвенирование*. Целевое секвенирование, как следует из названия, имеет меньшую исследовательскую мощность, чем WGS или WES, поскольку оно нацелено на определенные области гена и способно улавливать различные типы генетических вариаций от SNV до небольших делеций генов, дупликаций, вставок или перестроек генов, связанных с фенотипами заболеваний. Однако преимущества включают экономическую эффективность и управляемые данные для врачей, что упрощает принятие клинических решений с более конкретной информацией, относящейся к заболеванию. Оно может дать гораздо более глубокий охват до 5000× для редких аллелей при генетических заболеваниях, а также для малочисленных эволюционирующих мутантных клонов, возникающих в результате гетерогенности опухоли или эволюции заболевания при раке. Подход гена-кандидата или коммерчески доступные целевые панели являются результатом проектов WGS/WES, реализуемых в масштабе популяции. Зародышевую линию, а также соматические варианты можно протестировать с помощью целевых панелей NGS. Целевые панели работают на основе простого подхода обогащения путем амплификации с использованием пулов олигонуклеотидных праймеров, специфичных для региона. Библиотеки определенного размера, которые производятся, затем секвенируются и анализируются биоинформатически.

#### *Транскриптомика.* Секвенирование следующего поколения (NGS) оказало преобразующее влияние на транскриптомику, революционизировав нашу способность изучать транскриптом — полный набор молекул РНК в организме или конкретной популяции клеток. Технологии NGS предлагают высокопроизводительные и экономически эффективные методы профилирования и анализа молекул РНК, позволяя исследователям получить глубокое понимание экспрессии генов, альтернативного сплайсинга, регуляции некодирующей РНК и различных биологических процессов и заболеваний. Вот некоторые ключевые роли NGS в транскриптомике:

(а) Секвенирование мРНК (РНК-секвенирование): РНК-секвенирование является широко используемым приложением NGS в транскриптомике. Оно включает секвенирование и количественную оценку молекул мРНК, предоставляя комплексный снимок экспрессируемых генов в биологическом образце. Генерируя миллионы коротких прочтений секвенирования, NGS позволяет исследователям точно идентифицировать и количественно определять уровни экспрессии генов. Данные РНК-секвенирования можно анализировать для обнаружения дифференциальной экспрессии генов между различными условиями, обнаружения новых транскриптов, оценки альтернативных событий сплайсинга и изучения динамики экспрессии генов с течением времени или в различных тканях или типах клеток.

(б) Анализ альтернативного сплайсинга: Альтернативный сплайсинг, процесс, в котором один ген может генерировать несколько изоформ мРНК, вносит значительный вклад в сложность транскриптома. NGS дает возможность всесторонне изучать альтернативные паттерны сплайсинга. Выравнивая считывания РНК-секвенирования с референсным геномом, исследователи могут идентифицировать соединения сплайсинга и обнаруживать события альтернативного сплайсинга. Эта информация позволяет количественно определять и характеризовать изоформы транскриптов, предоставляя информацию о разнообразии изоформ, тканеспецифической экспрессии и функциональных последствиях альтернативного сплайсинга.

(с) Анализ длинных некодирующих РНК (lncRNA) и малых РНК: NGS облегчает изучение некодирующих РНК, которые играют важную роль в регуляции генов. Такие методы, как секвенирование малых РНК и секвенирование длинных некодирующих РНК, позволяют идентифицировать и характеризовать различные классы некодирующих РНК. Секвенирование малых РНК позволяет профилировать малые регуляторные РНК, включая микроРНК, пиРНК и snoRNA, предоставляя информацию об их роли в посттранскрипционной регуляции генов. Секвенирование длинных некодирующих РНК позволяет идентифицировать и анализировать длинные некодирующие транскрипты РНК, которые были вовлечены в различные биологические процессы и заболевания. Длинные считывания РНК-секвенирования могут информировать о связях между несколькими экзонами и выявлять вариации последовательностей (SNP) в транскрибируемой области. Секвенирование малых РНК — это нецелевой подход, позволяющий обнаруживать новые микроРНК и другие малые РНК. Транскриптом с исследованиями ChIP-seq в биологии рака помог понять новую роль некодируемых РНК, таких как sncRNA и lncRNA, в механизмах регуляции генов во время канцерогенеза/прогрессирования рака.

(г) Сборка и аннотация транскриптома: данные NGS могут быть использованы для реконструкции и аннотации транскриптома организма. Выравнивая считывания РНК-секвенирования с референтным геномом или используя подходы сборки de novo, исследователи могут идентифицировать новые транскрипты, варианты сплайсинга, нетранслируемые области и другие особенности транскрипта. Эта информация расширяет наше понимание сложности транскриптома и улучшает аннотацию референтных геномов, позволяя открывать ранее неизвестные гены и регуляторные элементы.

(е) Транскриптомика отдельных клеток: NGS способствовала появлению транскриптомики отдельных клеток, что позволило изучать профили экспрессии генов на уровне отдельных клеток. Технологии секвенирования РНК отдельных клеток (scRNA-Seq) позволяют профилировать транскриптомы отдельных клеток, предоставляя информацию о клеточной гетерогенности, идентификации типа клеток, анализе клеточной линии и динамике экспрессии генов в сложных тканях или процессах развития.

(ж) Интегративная транскриптомика: данные NGS из транскриптомики могут быть интегрированы с другими данными омики, такими как геномика, эпигеномика и протеомика, чтобы получить всестороннее понимание регуляции генов и биологических процессов. Интегративные подходы обеспечивают системный уровень молекулярного взаимодействия и позволяют идентифицировать ключевые регуляторные механизмы, лежащие в основе клеточных процессов и заболеваний.

#### *Эпигеномика.* Эпигеномика относится к изучению эпигенетических модификаций, которые являются наследственными изменениями в паттернах экспрессии генов, которые не включают изменения в последовательности ДНК. Наиболее распространенными типами изучаемых эпигенетических модификаций являются метилирование ДНК, модификация гистонов и метилирование РНК (эпитранскриптом). Эти химические метки, в свою очередь, изменяют доступность ДНК, ремоделирование хроматина и позиционирование нуклеосом. На эти модификации влияют факторы окружающей среды, такие как питательные вещества, загрязняющие вещества, токсичные вещества и воспаление. Знания и данные, полученные в результате полногеномного секвенирования у людей, растений и животных, помогли ученым лучше понять эти эпигенетические изменения, особенно метилирование и гидроксиметилирование ДНК. Эпигенетические изменения привлекли интерес исследователей и врачей к сложным расстройствам, таким как поведенческие расстройства, память, рак, аутоиммунные заболевания, зависимость, нейродегенеративные и психологические расстройства. Существуют различные платформы и анализы, разработанные для изучения эпигенетических модификаций, которые были очень хорошо описаны в других работах. NGS использовался для исследования эпигеномики, как обсуждается ниже:

(а) Профилирование метилирования ДНК: метилирование ДНК является важнейшей эпигенетической модификацией, которая играет решающую роль в регуляции генов и клеточных процессах. NGS позволяет проводить профилирование паттернов метилирования ДНК по всему геному с разрешением в один нуклеотид. Несколько стратегий, таких как бисульфитное секвенирование всего генома (WGBS) и бисульфитное секвенирование с пониженным представительством (RRBS), используют NGS для идентификации метилированных цитозинов. Однако RRBS основан на обогащении метилированных геномных регионов с использованием рестриктазного расщеплени. Эти методы позволяют исследователям изучать динамику метилирования ДНК, выявлять дифференциально метилированные регионы (DMR), связанные с заболеваниями, и понимать влияние метилирования на экспрессию генов.

(б) Картирование доступности хроматина: основанные на NGS методы, такие как анализ на транспозазо-доступный хроматин с использованием секвенирования (ATAC-seq) и DNase-seq, позволяют проводить профилирование доступности хроматина по всему геному. Эти методы идентифицируют области генома, которые доступны для ДНК-связывающих белков и факторов транскрипции, предоставляя информацию об элементах регуляции генов, усилителях и промоторах. Объединяя данные о доступности хроматина с другими эпигенетическими модификациями, данными об экспрессии генов и данными о связывании факторов транскрипции, исследователи могут раскрыть функциональные элементы в геноме.

(с) Анализ модификации гистонов: модификации гистонов, включая ацетилирование, метилирование, фосфорилирование и другие, являются критическими эпигенетическими метками, которые регулируют структуру хроматина и экспрессию генов. Иммунопреципитационное секвенирование хроматина (ChIP-seq) позволяет проводить профилирование модификаций гистонов по всему геному путем извлечения белка с помощью антител с последующим обогащением ДНК, связанной с белком, и секвенированием. Этот метод находит применение во многих различных областях исследований, таких как идентификация сайта связывания фактора транскрипции (TF), анализ модификации гистонов ДНК и метилирование ДНК. Для изучения модификаций гистонов антитела, нацеленные на модификации гистонов, используются для извлечения ДНК и секвенирования с использованием техники NGS. Полученные считывания выравниваются с референсным геномом, что позволяет идентифицировать паттерны модификации гистонов в определенных геномных областях. ChIP-Seq может дать представление об эпигенетической регуляции экспрессии генов, состояниях хроматина, а также об идентификации энхансеров и других регуляторных элементов.

(г) Анализ конформации хроматина: основанные на NGS методы, такие как Hi-C и 4C-seq, позволяют исследовать трехмерную организацию и взаимодействия хроматина. Эти методы фиксируют дальние взаимодействия хроматина и позволяют строить карты взаимодействия хроматина. Интегрируя данные о трехмерной конформации хроматина с эпигенетическими модификациями, данными об экспрессии генов и функциональными аннотациями, исследователи могут получить представление о пространственной организации генома и понять, как она влияет на регуляцию генов.

(е) В дополнение к этим автономным подходам данные NGS из эпигеномики могут быть интегрированы с данными транскриптомики для раскрытия связи между эпигенетическими модификациями и экспрессией генов. Интеграция профилей метилирования ДНК с данными РНК-секвенирования может идентифицировать дифференциально метилированные регионы (DMR), связанные с изменениями экспрессии генов. Интеграция данных о модификации гистонов и доступности хроматина с РНК-секвенированием позволяет идентифицировать регуляторные элементы, связанные с определенными паттернами экспрессии генов, и исследовать эпигенетические регуляторные механизмы.

#### *Метагеномика.* Метагеномика занимается прямым генетическим анализом прокариотического генома, включая бактерии, грибы и вирусы, содержащиеся в образце, либо с помощью целевого подхода, либо с помощью подхода ПЦР с лигированием адаптеров для секвенирования методом дробовика в культурально-независимой манере. Гипервариабельная область в генах рибосомальной РНК 16S или 18S бактерий и грибов используется в целевом подходе. Смесь консервативных и гипервариабельных областей помогает в идентификации каждого вида бактерий из образца. Аналогично, для идентификации видов грибов для амплификации выбираются области ITS1 и ITS2, охватывающие ген 5.8S рРНК генома гриба. Для секвенирования вирусного генома прочтения, полученные с помощью NGS (метод дробовика), снова являются культурально-независимым методом изучения вирусного разнообразия, распространенности и функционального потенциала вирусов в окружающей среде. Все отфильтрованные прочтения сопоставляются с референсной последовательностью человека, а оставшиеся, неотображенные прочтения сопоставляются с вирусной геномной базой данных NCBI RefSeq. Также доступны целевые панели вирусного и бактериального генома, например, ChapterDx для обнаружения HR HPV и микробной инфекции, панель устойчивости к препаратам ВИЧ, панель AMR, панель желудочно-кишечных расстройств.

На основе сходства нуклеотидных последовательностей предварительно обработанные последовательности кластеризуются с 97% сходством в операционные таксономические единицы (OTU). OTU сравниваются с базой данных для идентификации микроорганизмов. Для анализа прочтений ампликона 16S используется несколько аналитических конвейеров. Для образцов метагеномики дробовика таксономические и функциональные профили могут быть получены различными подходами. Секвенирование микробиома может идентифицировать полный спектр микробных видов, присутствующих в образце. Результаты являются высококоличественными, и можно изучать бактериальные сообщества в определенном интервале условий. Платформа NGS также может генерировать прочтения для видов с низким содержанием в образце.

1. **Автоматизация процесса секвенирования.** Автоматизация процесса секвенирования играет ключевую роль в современном геномном анализе, позволяя значительно ускорить получение генетической информации и сократить затраты. Автоматизированные системы не только повышают скорость и точность секвенирования, но и открывают новые возможности для исследований, диагностики и персонализированной медицины. С развитием технологий секвенирования нового поколения (NGS) и улучшением обработки данных процесс автоматизации выходит на новый уровень, удовлетворяя запросы как фундаментальной науки, так и клинической практики.

**Историческая перспектива автоматизации секвенирования**

Первые методы секвенирования, такие как метод Сэнгера, требовали ручного труда и были длительными. В 1970-х годах секвенирование производилось поэтапно, требовало значительных временных затрат и человеческих ресурсов. Несмотря на свою надежность, метод Сэнгера оказался ограничен в масштабах, особенно в проекте "Геном человека", который потребовал автоматизации. В 2000-х годах начали разрабатываться новые автоматизированные методы, известные как секвенирование нового поколения (NGS), которые привели к значительному сокращению стоимости секвенирования одного генома и сделали возможным массовое секвенирование ДНК.

**Основные технологии автоматизированного секвенирования**

Наиболее распространенными технологиями автоматизированного секвенирования являются платформы, основанные на NGS, такие как Illumina, Ion Torrent и PacBio. Эти платформы значительно ускорили получение геномных данных и упростили процесс анализа.

1. **Illumina**: Основано на использовании реверсивных терминаторов, позволяющих синтезировать новые цепи ДНК с последовательным добавлением меток к каждому нуклеотиду. Автоматизация процесса позволяет одновременно секвенировать миллиарды фрагментов ДНК, что делает платформу одной из самых эффективных для масштабного геномного анализа. Illumina успешно применяется для секвенирования экзома, РНК и целых геномов с высокой точностью.
2. **Ion Torrent**: Использует метод измерения изменения рН для определения нуклеотидов в процессе синтеза новой цепи ДНК. Эта технология не требует флуоресцентной маркировки и оптики, что делает её менее затратной. Ion Torrent автоматизировал процесс секвенирования, обеспечив более доступное и быстрое получение данных, что особенно полезно в клинических лабораториях и для анализа патогенных микроорганизмов.
3. **PacBio и Oxford Nanopore**: Длинночитающие технологии, позволяющие получать последовательности в десятки тысяч пар оснований. Эти системы значительно улучшили качество сборки геномов, позволив преодолеть трудности, связанные с длинными повторяющимися элементами ДНК. Платформы PacBio и Oxford Nanopore активно используются в геномике сложных организмов, метагеномике и эволюционных исследованиях.

**Программное обеспечение и алгоритмы для автоматизации**

С развитием автоматизации возросла необходимость в разработке эффективных алгоритмов и программного обеспечения для обработки больших объемов данных, получаемых при секвенировании. Одним из примеров является использование платформы Galaxy, обеспечивающей интегрированное рабочее пространство для анализа данных. Такие пайплайны, как nf-core, делают возможной стандартизацию и автоматизацию сложных процессов анализа, таких как фильтрация, выравнивание, сборка и аннотация геномных данных.

Современные алгоритмы машинного обучения также начинают играть значительную роль в автоматизации секвенирования. Алгоритмы на основе глубокого обучения помогают улучшить точность предсказания последовательностей и корректировать ошибки, возникающие при работе с длинночитающими технологиями. Программное обеспечение, основанное на машинном обучении, используется для идентификации генов, предсказания структурных вариаций и обнаружения мутаций.

**Автоматизация секвенирования в клинической диагностике**

Одной из ключевых областей применения автоматизированного секвенирования является клиническая диагностика. Автоматизация позволяет с высокой точностью выявлять мутации, связанные с наследственными заболеваниями, проводить тестирование на предрасположенность к раку, инфекциям и определять генетические вариации, влияющие на ответ на лекарства. Например, платформы, такие как Illumina и Ion Torrent, уже используются для секвенирования экзомов пациентов, что позволяет идентифицировать клинически значимые мутации в пределах 1-2 дней.

Автоматизированные системы секвенирования играют важную роль в вирусологии и эпидемиологии. Во время пандемии COVID-19 автоматизация секвенирования геномов вируса SARS-CoV-2 помогла отслеживать эволюцию и распространение вируса, позволяя быстро выявлять новые штаммы и контролировать эффективность вакцин. Доступность автоматизированных систем также способствовала разработке точных тестов для диагностики вирусных инфекций.

**Текущие вызовы и перспективы развития автоматизации секвенирования**

Несмотря на значительный прогресс, автоматизация секвенирования сталкивается с рядом вызовов. Большие объемы данных, создаваемые NGS и длинночитающими платформами, требуют значительных вычислительных мощностей и надежных систем хранения данных. Обработка и анализ таких данных продолжают оставаться сложной задачей, особенно в случаях, когда требуется высокий уровень точности и воспроизводимости.

Кроме того, стандартизация автоматизированных пайплайнов для обработки геномных данных имеет важное значение для повышения точности и согласованности результатов между лабораториями. Введение стандартов, таких как VCF (Variant Call Format) и BAM (Binary Alignment Map), позволило унифицировать хранение и обмен данными, но еще предстоит разработать универсальные стандарты для автоматизированной интеграции данных в клинические системы.

Будущее автоматизации секвенирования связано с развитием более интегрированных систем, которые будут включать в себя не только секвенирование, но и автоматизированный анализ и интерпретацию данных. Такие системы могут стать частью клинических лабораторий, предлагая врачам персонализированные данные о пациентах в течение нескольких часов после поступления образца. Это будет особенно полезно в онкологии, генетической диагностике и инфекционных заболеваниях, где время имеет решающее значение.

Автоматизация процесса секвенирования произвела революцию в биологии и медицине, позволив ученым и врачам получить доступ к геномной информации быстрее и эффективнее, чем когда-либо прежде. Современные технологии NGS, длинночитающие платформы и передовые алгоритмы обработки данных делают возможным исследование генетического материала на совершенно новом уровне. В ближайшем будущем автоматизация секвенирования продолжит развиваться, открывая новые возможности для диагностики, лечения и научных исследований.

**Вопросы для самоконтроля:**

1. Какие основные преимущества и недостатки каждого из методов секвенирования (например, Illumina, PacBio, Oxford Nanopore) при анализе сложных геномов и как это влияет на выбор технологии для конкретного исследования?
2. Как автоматизация процесса секвенирования повлияла на точность и скорость обнаружения редких генетических вариаций, особенно в клинической диагностике?
3. Какие уникальные технические и алгоритмические проблемы возникают при автоматизации обработки данных длинночитающих технологий, таких как PacBio и Nanopore?
4. Как стандартизация форматов хранения и обработки данных (такие как VCF и BAM) способствуют автоматизации и воспроизводимости в клинических и исследовательских лабораториях?
5. Какие алгоритмы и методы машинного обучения играют решающую роль в обработке и интерпретации больших объемов данных секвенирования, и какие проблемы они помогают решать?
6. В каких направлениях автоматизация секвенирования и анализа геномов может улучшить диагностику и лечение заболеваний, особенно в персонализированной медицине?
7. Какую роль играют современные технологии облачных вычислений и распределенной обработки данных в повышении эффективности автоматизированных систем секвенирования и анализа геномов?